IDENTIFICATION OF NERVE GROWTH FACTOR (NGF) RECEPTOR MODULATORS FOR NERVE REGENERATION

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENT OF FOR THE DEGREE OF **BACHELOR OF TECHNOLOGY**

IN

BIOTECHNOLOGY

Project Submitted by:-

ANAMIKA YADAV

111BT0497

Project Supervised by:-

Prof. Sirsendu Sekhar Ray

Department of Biotechnology and Medical Engineering

DEPARTMENT OF BIOTECHNOLOGY AND MEDICAL ENGINEERING

NIT ROURKELA

2014-2015

NATIONAL INSTITUTE OF TECHNOLOGY, ROURKELA

CERTIFICATE

This is to certify that thesis entitled "IDENTIFICATION OF NERVE GROWTH FACTOR RECEPTOR MODEULATORS FOR NERVE REGENERATION" submitted by **Ms. Anamika Yadav** to the National Institute of Technology, Rourkela for the partial fulfillment of the requirements for the degree of Bachelor of Technology in BIOTECHNOLOGY contains the bonafide work carried out by her in the Department of Biotechnology and Medical Engineering under the supervision and guidance of the undersigned. To the best of my knowledge, the thesis or any part of it has not been submitted earlier to any other University/Institute for the award of any degree or diploma.

DATE: 11.05.2015

PLACE: Rourkela

Dr. Sirsendu Sekhar Ray

Assistant Professor

Department of Biotechnology and Medical Engineering

National Institute of Technology, Rourkela

ACKNOWLEDGEMENT

I would like to offer my sincere gratitude and warmest regard to my guide Dr. Sirsendu Sekhar Ray, Department of Biotechnology and Medical Engineering, NIT Rourkela, for his valuable guidance and encouragement. He has been an unwavering source of knowledge and inspiration throughout my research work. I sincerely thank him for the time and patience he devoted for this work.

I am also grateful to my Head of Department, Dr. Krishna Pramanik, and the entire faculty for their cooperation and motivation. I would also like to extend my gratitude for the facilities provided in the Bioinformartics Laboratory, Department of Biotechnology and Medical Engineering.

I would also like to extend a special note of thanks to Ms. Priyanka Goyal, M.Tech Final Year, and Mr. Praveen Kumar, M.Tech Final Year, who helped me and guided me through the difficulties I faced regarding this project. I would also like to thank Ms. S. Meghasmita, Ms. Samapika Mishra and all my friends for their help and constant support.

Anamika Yadav

ABSTRACT

The physiological role of the neurotrophin nerve growth factor (NGF) has been characterized, since its discovery in the 1950s, first in the sensory and autonomic nervous system, then in central nervous, endocrine and immune systems. The biological function of the NGF is the maintenance and survival of the peripheral and central nervous systems, which makes them of great therapeutic interest for the treatment of a number of neurodegenerative diseases. Identifying the ligands which can enhance the effect of this growth factor can be very vital for using it as a pharmacological tool in the treatment of various CNS diseases like the Alzheimer's Disease, Parkinson's Disease, Glaucoma and advanced optic nerve atrophy and various peripheral neuropathies. The property of the NGF to promote neuronal regeneration can also be applied in neural tissue engineering applications. Molecular docking was performed by Autodock4.2 using Lamarckian Genetic Algorithm to identify the ligands with high binding affinity with the NGF receptor TrkA. The best docked ligands were further analyzed for their potential as drug candidates by predicting their ADME properties using online PreADMET tool. From the combined analysis of molecular docking and ADME properties, a group of ligands having similar structure were further investigated to reveal that their similar structure led to their similar binding affinities with the NGFR. Quantitative Structure-Analysis Relationship (QSAR) models were studied using oleanolic acid as base structure, and 2D and 3D QSAR mathematical models were generated by using VlifeMDS (Version 4.3) to predict the activities of new analogs. These models can also be used to design new analogs with better activities by modulating the structure based on the results obtained.

Key words: Nerve Growth Factor, neuronal regeneration, docking, ADME, QSAR

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1:

INTRODUCTION

1.1 NEURAL TISSUE ENGINEERING

Neural tissue engineering is aimed at promoting nerve regeneration and repairing the damage caused to neurons of both the Peripheral Nervous System (PNS) and the Central Nervous System (CNS). Loss of function or injuries to neurons in the CNS and PNS can be caused due to strokes, trauma or neurodegenerative disorders. In the PNS, if the injuries are small the nerves can regenerate on their own, but large injuries require surgical treatments often involving grafts [1]. In CNS, nerves fail to regenerate on their own as the native cellular environment of the CNS inhibits regeneration of cells [2]. Thus, due to the difficulty of nerve cells to repair or regenerate on their own, neural tissue engineering has emerged as a promising attempt at nerve repair. Various approaches to neural tissue engineering have been adopted in the past like physical guidance of neurites (axons or dendrites), use of grafts and biomaterials [3]. Lately, the role played by neurotrophic factors in neural regeneration is being extensively studied, and the mechanisms by which neurotrophic factors promote the survival and process outgrowths of neurons have also been well characterized [4].

1.2 NEURODEGENERATIVE DISORDERS

Neurodegenerative Disorders refer to a range of conditions that alter the structure, and also the function of neurons in the human brain. They include diseases like Alzheimer's Disease (AD), Huntington's Disease (HD), Glaucoma, Amyotrophic Lateral Sclerosis (ALS), Parkinson's Disease (PD), etc. Most of these diseases are incurable as the neurons once damaged cannot repair or replace themselves. In AD, there occurs a selective loss of neurons in the cerebral cortex and certain regions below the cortex. It is characterized by extracellular amyloid plaques and intracellular tau protein tangles [5]. In PD, there is a loss of neurons present in the substantia nigra region notably altering the central component of the pars compacta. Mutations in the parkin protein produce a parkinsonian syndrome responsible for some genetic forms of PD. The innate features of PD are rigidity, rest tremor and bradykinesia [6-7].

1.3 NERVE GROWTH FACTOR AND ITS RECEPTOR

Nerve Growth Factor (NGF), the earliest neurotrophic factor to be discovered, is the best-characterized member of the neurotrophin family. It was discovered almost fifty years ago by

Rita Levi Montalcini. The neurotrophin family is a group of functionally and structurally similar homodimeric proteins, that include NGF, brain-derived neurotrophic factor (BDNF), neurotrophin-3 (NT-3), neurotrophin-4/5 (NT-4/5), and neurotrophin-6 [8-11]. The central biological function of NGF is to facilitate the maintenance and survival of the neurons of the CNS and PNS, which instills great interest in them for treating neurodegenerative disorders [11-12].

The structure of NGF monomer has two pairs of antiparallel β-strands that twist around each other to form the central part of an elongated molecule. On one end, there are three hairpin loops (L1,L2 and L4 in figure 1), and on the other there is a cysteine knot motif that secures the folds of the strands and stabilizes the molecule in this conformation. It's biologically active form has two monomers that are arranged parallel to each other, forming a closely packed homodimer [12].

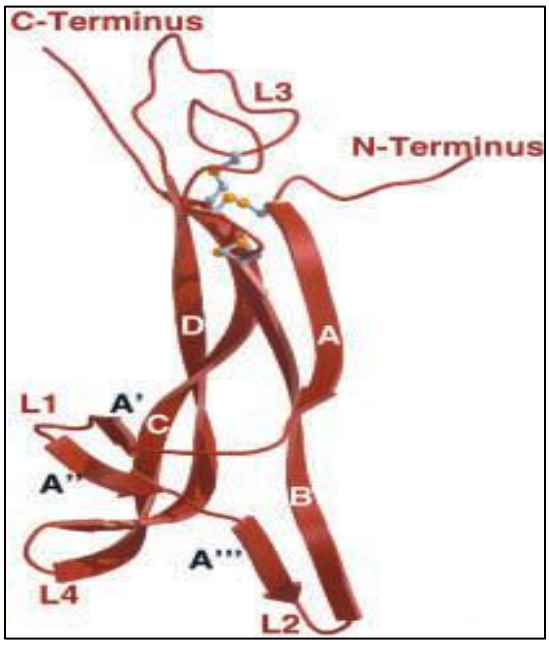

Figure 1: Ribbon structure of NGF monomer

NGF is known to control actively the synthesis of neuropeptides and neurotransmitters in sympathetic and sensory neurons. The cholinergic neurons in the basal forebrain complex (BFC) are highly dependent on NGF for their survival and basic functions like motivation, attention, arousal, consciousness, and memory. As BFC neurons are compromised in

Alzheimer's disease (AD), NGF is considered to be a potential therapeutic factor for such neurodegenerative disorders [13-14].

The neurotrophins exercise their action by binding to two groups of receptors - the p75 neurotrophin receptor, also known as the low-affinity neurotrophin receptor (LNGFR) and the Trk receptors. p75 binds to all the neurotrophins with almost the same affinity. There are three Trk receptors namely; TrkA which preferably binds to the NGF, TrkB which preferably binds to BDNF and NT-4/5 and TrkC which at binds to NT-3 [15-18]. The Trk receptors (receptor tyrosine kinases) dimerize to their extracellular portions upon binding to neurotrophins. This dimerization event causes several tyrosine residues present in the cytoplasmic domain of the receptors to undergo autophosphorylation, thereby activating the intracellular signal transduction pathway. The TrkA receptor interacts with NGF independently and mediates NGF signaling [19]. Thus to regulate NGF activity, the TrkA receptor (NGFR) can be modulated by identifying ligands that can act as NGF stimulators. Computer Aided Drug Discovery (CADD) can be used to identify active drug candidates and optimize lead molecules.

1.4 OBJECTIVE

To identify modulators of NGFR (TrkA) for neuronal regeneration in neurodegenerative disorders and for tissue engineering applications.

1.4.1 To prepare a database of phytochemicals

1.4.2 To prepare protein target

1.4.3 To perform molecular docking of all the phytochemicals with the target protein for the identifying novel drug candidates

1.4.4 To study ADME properties of the best docked ligands

1.4.5 To generate and validate QSAR models

1.5 PLAN OF WORK

Plan of Work	7th Semester		8th Semester	
	Mid Semester	End Semester	Mid Semester	End Semester
Literature Review				
Identification of Phytochemicals				
Preparation of target protein and ligands for docking				
Molecular Docking				
ADMET Predictions				
QSAR Studies				
Manuscript writing				

CHAPTER 2:

LITERATURE REVIEW

2.1 NEURODEGENERATIVE DISORDERS

2.1.1 Alzheimer's Disease

Alzheimer's disease is a neurodegenerative disorder that leads to decline in memory, language, problem-solving and other cognitive skills of the affected person. This happens because the neurons of the regions involved in cognitive skills like the cerebral cortex and subcortical regions have been damaged and lose their normal function. The neuronal damage eventually affects other parts of the brain responsible for basic bodily functions. In the later stages, AD progresses rapidly and is ultimately fatal. According to the Alzheimer's Association Report published in 2015, an estimated 5.3 million Americans are affected by AD; of which 5.1 million are above the age of 65 years, and approximately 200,000 are below the age of 65 years. In 2013, official death certificates recorded 84,767 deaths due to AD, making it the sixth leading cause of deaths in the United States [20].

In AD, the neurons of the Basal forebrain complex and the cerebral cortex lose their function or/and structure. AD is characterized by the occurrence of two aggregates – extracellular amyloid plaques and intracellular neurofibrillary tangles of hyperphosphorylated tau protein. The amyloid plaques are found to contain small toxic cleavage products (Aβ40 and Aβ42) of the amyloid precursor protein (APP). Many mechanisms for the neurodegeneration pathway in AD have been proposed. According to one of the suggested views, Aβ protofibrils stimulate microglia, triggering an inflammatory response and production of neurotoxic cytokines. Another mechanism suggests that tau protein tangles and amyloid aggregates relay adverse effects or harmful responses to neurons by disabling the dendritic and axonal transport. Another mechanism indicates Aβ injury is due to synaptic dysfunction and loss. Synaptic dysfunction is believed to contribute to loss in memory and other cognitive deficiencies in AD. Aβ may also cause undesired effects by binding to redox-reactive metals which lead to release of free radicals. Oxidative stress from mitochondrial dysfunction, caused by the blocking of respiratory complex I have also been indicated as a potential pathway [21-24].

2.1.2 Parkinson's Disease

Parkinson's Disease is another severe neurodegenerative disorder that affects a person's movements. In PD, the dopamine-producing neurons of the substantia nigra region are compromised as a result of which initially movement-related activities are impaired and gradually cognitive. The marker of PD is characterized by loss of neurons within the substantia nigra region that affect the central component of the pars compacta. α-synuclein-immunoreactive inclusions known as Lewy bodies are the characteristic lesions of PD [25]. Although ∝-synuclein was the first protein to be linked with familial PD, eventually various additional genetic loci were determined which can now be linked to the development of PD. For example, mutations and polymorphisms in *UHCL1*, the gene encoding ubiquitin carboxy-terminal hydrolase L1 (UCH-L1), are linked to PD. Also, autosomal mutations in *PARK2*, gene that encodes parkin protein, are the cause for up to half of the juvenile and early-onset parkinsonism [26].

2.2 ROLE OF NGF IN VARIOUS DISEASES

NGF is produced in large volumes in the cortex, the pituitary gland, and the hippocampus. It is also produced in certain other areas like basal ganglia, spinal cord, thalamus and the retina. As we now know, NGF has a significant role to play in the survival, maintenance and functioning of neurons in PNS and the cholinergic neurons in the CNS. Given this critical role played by NGF, a large number of diseases (especially neuropathies and degenerative disorders) are associated with altered NGF levels or/and changed expression of its receptors.

The cholinergic neurons of the basal forebrain complex (BFC), which are highly affected in the AD depend on NGF for their survival and proper functioning. It has been found that AD patients show reduced levels of NGF in BFC of their brain. It was concluded from rat models, that NGF administration reduced cholinergic neuron death and cholinergic atrophy in rats [27-28]. It is also now accepted based on experimentation, that NGF can act on the two characteristic hallmarks of AD - β-amyloid neurotoxicity and tau hyperphosphorylation. In Down's Syndrome, another neurodegenerative disorder, a similar reduction in NGF levels in the BFC, like in AD, is observed [29]. Another disease called glaucoma, which is the most common cause of blindness, is marked by the degeneration of the retinal ganglion cell (RGC)

and optic nerve atrophy. Glaucoma can be considered a neurodegenerative disease as the RGCs are a type of specialized neurons that pick up visual information from photoreceptors and convey the signal to the brain. In glaucoma, the increased intraocular pressure reduces the NGF levels in the cerebrospinal fluid and lateral geniculate nucleus. Thus using NGF as a means to attenuate RGC degeneration is a potential strategy [30]. NGF can also be exploited as a therapeutic tool in the treatment of other neurodegenerative disorders like PD, ALS, and Huntington Disease etc.

Apart from the neurodegenerative disorders, NGF is known to have many other clinical uses in other diseases. It has been observed from human pathologies and various animal models, that disease-associated peripheral neuropathies, like diabetes and HIV, could be linked to irregularities in the regulation of synthesis, transport and usage of NGF by PNS neurons. For example, diabetes is characterized by degeneration and loss of function of some types of PNS neurons and deficits in NGF transport [31]. NGF administration in animal models of diabetic neuropathies has shown to reverse the neuropathic symptoms by preserving the compromised PNS neurons and restoring their normal activity [32]. NGF also has an important role to play in the response to neuronal injury. Peripheral nerve injury is often followed by remarkable neuronal cell death in dorsal root ganglion (DRG). NGF controls the synthesis of neuropeptides and neurotransmitters in sensory neurons. This ability of NGF helps it to counteract the effect of nerve injury on the DRG cells. NGF administration has been observed to sustain axonal regeneration. NGF also acts as a regulatory factor for several non-neuronal cells which express NGF receptors on their surface. NGF receptors are expressed in immune cells allowing NGF to control cell differentiation and modulating the immune response. NGF influences differentiation, survival and phenotypic features of hematopoietic stem cells, granulocytes, monocytes and lymphocytes. NGF is also produced and utilized in skin cells like keratinocytes, and NGF deregulation has been observed in skin ulcers [33].

2.3 DATABASE OF PHYTOCHEMICALS

Given the important role of NGF in various diseases, it can be considered as an efficient potential treatment strategy for the disorders mentioned above. Till date, NGF administration through intracerebroventricular (ICV) injections and NGF-based gene therapy have been used for the delivery of NGF to the affected sites. One another strategy can be to stimulate

the low levels of NGF in the affected areas to prompt regeneration and reinstate normal function of neurons by using external factors like ligands that can bind to the NGFR. For this, a database of phytochemicals with neuroprotective properties was created in this study using literature survey. Phytochemicals (ligands) that turn out to have a high binding affinity to the NGFR can be in turn used as pharmacological tools to promote NGF activity in various diseases.

Table 1: Database of phytochemicals (first few of the 235 phytochemicals under study)

PLANT	PHYTOCHEMICALS	SOURCE
1. Turmeric	curcumin, cineol, α-phellanderene, sabinene, curlone, ar-curcumene, a-phellandrene, b-caryophyllene, b-sesquiphellandrene, borneol, b-bisabolene, sesquiter, d-3-carene, zingiberene	[A Touch of Turmeric: Examining an Ayurvedic Treasure: Prianca Madi Reddi et al]
2. Brahmi (bacopa monnieri)	brahmine, d-mannitol, herpestine, hersaponin, beta-sitosterol, monnierin, stigmastarol, betulic acid	[Neuroprotective potential of phytochemicals: G. Phani Kumar and Farhath Khanum]
3. Centella asiatica	asiaticoside, betulic acid, oxyasiaticoside, isobrahmic acid, centelloside, brahmic acid, brahmoside, madecassic acid, brahminoside, asiatic acid, thankunoside, isothankunoside	[Neuroprotective potential of phytochemicals: G. Phani Kumar and Farhath Khanum]

4.Withania somnifera	cuscohygrine, scopoletin, tropine, anahygrine, withasomnine, anaferine, withanine, somnine, isopelletierine, withananine, withananinine, somniferinine, pseudo-withanine, withaferine, somniferine, condensed tannins, visamine	[Phytochemical and Pharmacological Profile of Withania somnifera Dunal -A Review: Qamar Uddin et al]
5. Acorus calamus	sesquiterpenes, flavonoids, α- and β-asarone, calamen, clamenol, calameon, acorine, eugenol, pinene, camphene,elemicine, calamendiol, cisisoelemicine, cis and trans isoeugenol, P-cymene, bgurjunene, α-selinene, β-cadinene, terpinen-4-ol, camphor, calacorene, linalool, aterpineol, acorone, spathulenol, acrenone, acoragermacrone, 2-deca-4,7 dienol, Acoradin, shyobunones, preisocalamendiol, 2, 4, 5- trimethoxy benzaldehyde, galangin, 2, 5 dimethoxy benzoquinone.	[Sweet flag (Acorus calamus Linn.): An incredible medicinal herb, Hashmat Imam et al]

CHAPTER 3:

METHODS AND MATERIALS

3.1 BIOINFORMATICS TOOLS AND SOFTWARES

1. National Centre for Biotechnology Information (NCBI)
2. Protein Data Bank (PDB)
3. UniProt
4. PubChem
5. PRODRG Server
6. ArgusLab
7. UCSF Chimera
8. AutoDock 4.2
9. Cygwin
10. PreADMET
11. ChemDraw
12. VlifeMDS 4.3
    - 12.1 2D QSAR Module
    - 12.2 Molecular Alignment Module
    - 12.3 3D QSAR Module

3.2 PROTEIN STRUCTURE RETRIEVAL

The three dimensional crystal structure of the target protein (NGFR) was obtained from an online database- Protein Data Bank (http://www.rcsb.org/pdb/). To predict the active site of the protein and perform further docking studies, miscellaneous ligands were removed to obtain the isolated structure of the protein using Argus Lab Software. Other heteroatoms such as water, ions, etc. present in the protein model were removed too. The geometry and energy minimization of the protein were performed, and the optimized protein was saved in .pdb format.

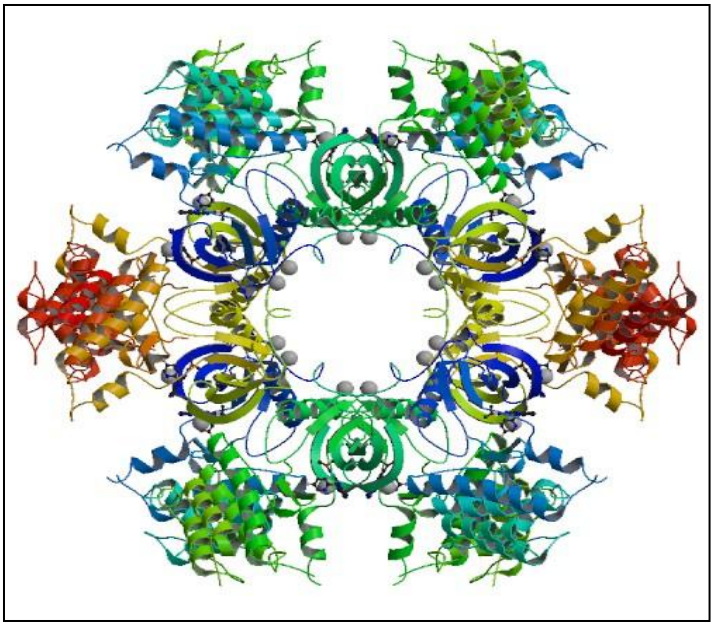

Figure 2: Ribbon Structure of TrkA receptor (PDB ID:4AOJ)

3.3 PROTEIN ACTIVE SITE PREDICTIONS:

The active sites of target protein were identified from the literature on UniProt (http://www.uniprot.org/). The binding and active sites present in TrkA receptor were found to be Asp-399, Tyr-496, Lys-544, Asp-650, and Tyr-791. Asp-650, being the active site, was used as the center residue for grid box.

3.4 PREPARATION OF DATABASE OF PHYTOCHEMICALS (LIGANDS):

A database of plants with neuroprotective properties and their phytochemicals was prepared. A total of 235 phytochemicals from 26 plants were collected and their structures downloaded. The three dimensional structures of these phytochemicals were retrieved from PubChem database (http://pubchem.ncbi.nlm.nih.gov/). The .sdf files thus obtained were converted to PDB structure of the phytochemicals by using the online PRODRG Server (http://davapc1.bioch.dundee.ac.uk/cgi-bin/prodrg/run.html#DRGPOH.PDB).

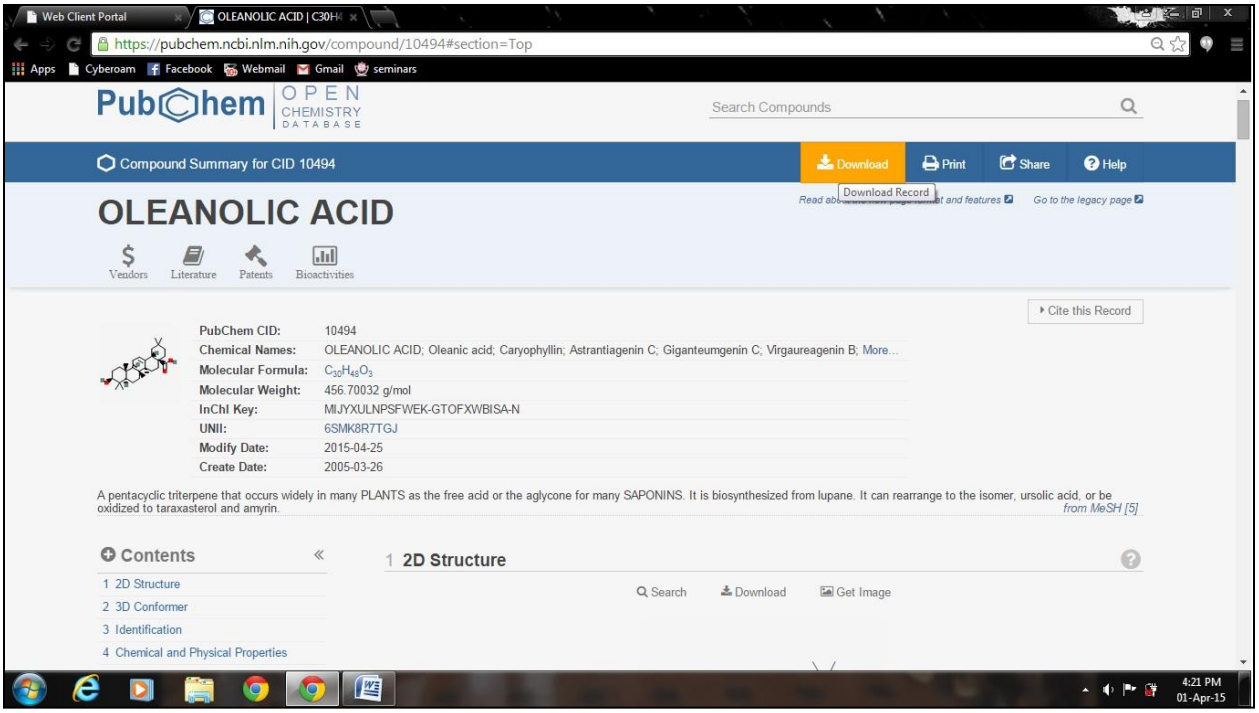

Figure 3: Screenshot of oleanolic acid in PubChem

3.5 MOLECULAR DOCKING:

Docking is an efficient technique to predict the binding affinity and orientation of various ligands to a particular protein. Thus, the molecular docking of the NGF receptor TrkA (protein) and phytochemicals (ligands) was performed at the active site of the target protein using the AutoDock4.2 software (version 1.5.6). In Autodock, polar hydrogens were added to the protein model and Kollman charges were assigned to the protein structure. The ligand was added, and both the protein and ligand structures were saved in .pdbqt formats. The grid maps representing the proteins were calculated using autogrid and grid size was set to 60*60*60 points, and it was fixed on the active site Asp-650. The resultant file was saved in .gpf format. Docking of ligand was carried out using Lamarckian Genetic Algorithm as the search function, and the final structure was saved as .dpf format.

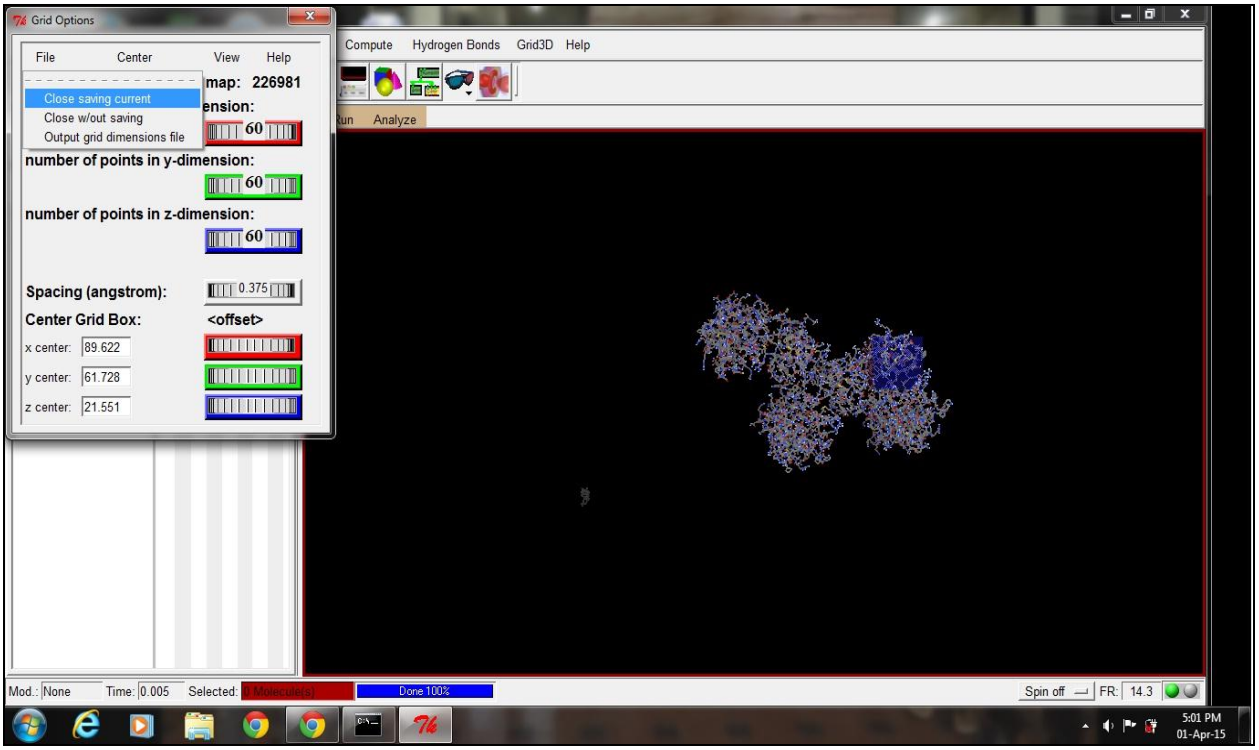

Figure 4: Screenshot of fixing the grid on oleanolic acid while Docking in AutoDock

3.6 CONVERSION OF .gpf AND .dpf FILES TO PDB FILES USING CYGWIN

The .gpf and .dpf files generated from Autodock were then converted to .glg and .dlg files using the Cygwin command prompt. These files were further converted into PDB formats. These PDB formats are the docked structure of the protein and ligand which can be viewed in ArgusLab or Chimera. The .dlg file can be opened in a WordPad and the lowest binding energy and estimated inhibition constant of the ligands were tabulated. Lowest binding energy will signify the highest binding affinity to the target protein.

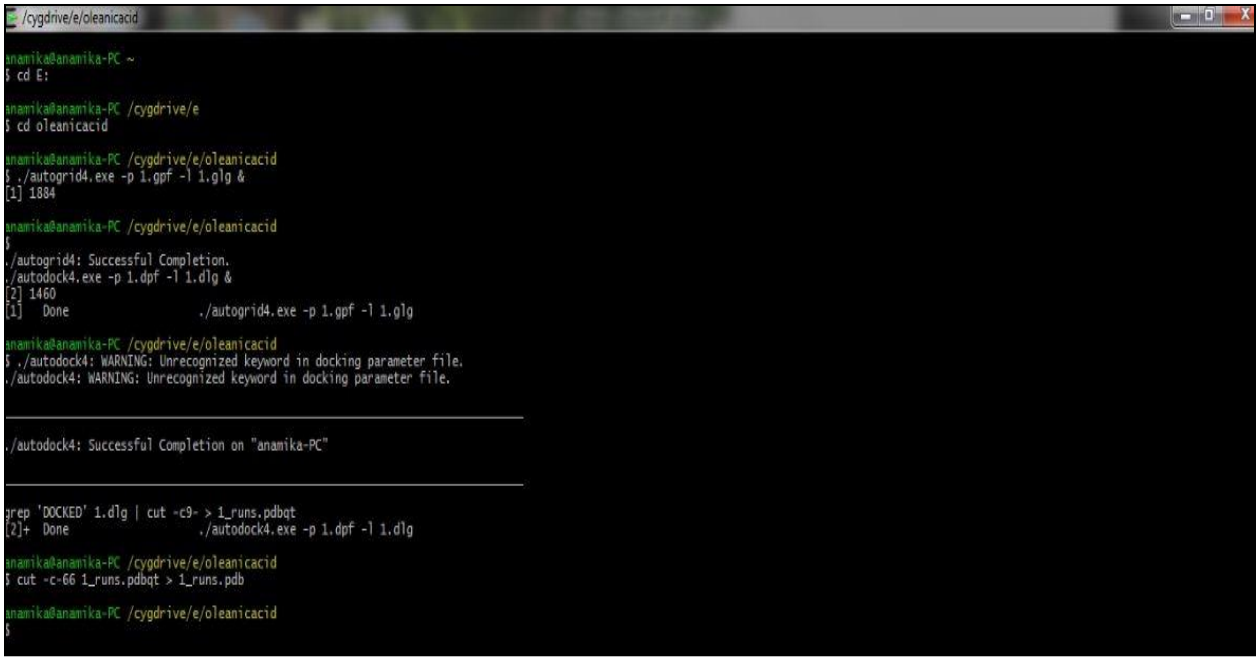

Figure 5: Screenshot of Cygwin command prompt

3.7 ADME PREDICTIONS:

A very important challenge in the drug discovery process is that more than half of the drugs in a drug discovery process fail to commercialize because of deficits in ADME and toxicity properties. Hence, early prediction of these properties are required these days to increase the success rate of lead molecules. The ligands having a good binding affinity with the NGFR were further examined for their ADME properties and toxicity as part of lead optimization. This was done using an online tool PreADMET (http://preadmet.bmdrc.org/). The structure of the ligand was submitted in the ADME worksheet, and the output file was saved.

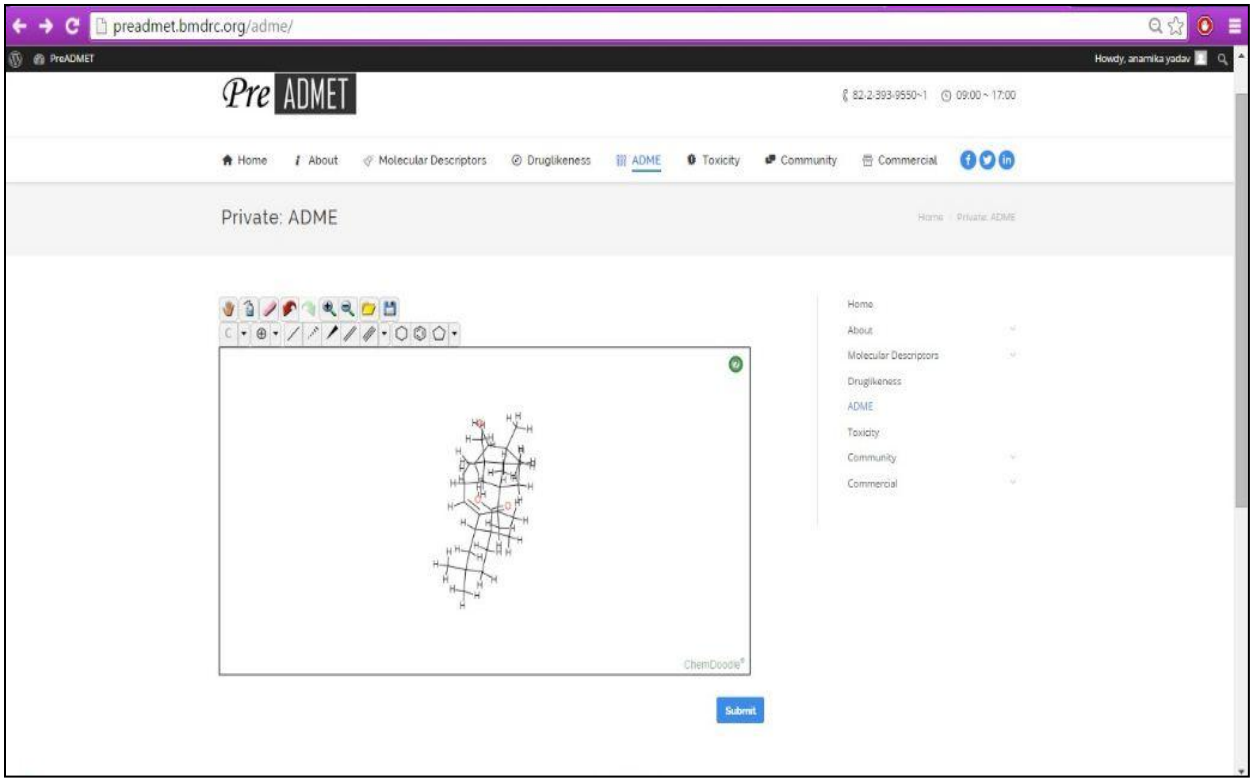

Figure 6: Screenshot of PreADMET tool

3.8 QUANTITATIVE STRUCTURE-ACTIVITY RELATIONSHIP (QSAR):

QSAR is a method to understand the relationship between a molecule's bioactivity and its structure. The bioactivity of a molecule is directly dependent on its structure, and thus molecules having similar structure tend to have similar activities. QSAR can, therefore, be used to design a mathematical model based on a group of analogs of a lead molecule that can predict the activity of other similar molecules.

2D and 3D QSAR were performed using the VLife Molecular Design Suite (VLifeMDS) version 4.3. Thirty substituted oleanolic acid derivatives were taken and their activity data were given as the negative logarithm of inhibition constant (-log K_i). These activity data are used as the dependent variable for our QSAR studies.

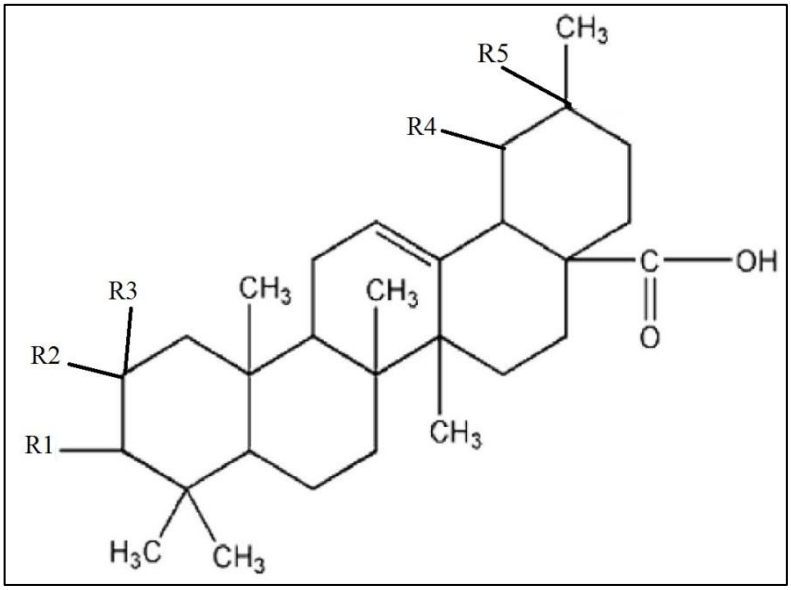

Figure 7: Skeleton structure used for QSAR (oleanolic acid)

Table 2: Oleanolic Acid derivatives with their corresponding activity

S.NO.	**R1**	**R2**	**R3**	**R4**	**R5**	**B.E. (kcal/mol)**	**Ki**	**ACTIVITY (-log Ki)**
OA1	OH	H	H	H	CH_3	-11	8.61 nM	8.064
OA2	OH	H	H	OH	CH_3	-11.42	4.28 nM	8.368
OA3	OAc	H	H	OH	CH_3	-14.39	28.18 pM	10.55
OA4	OH	=O		OH	CH_3	-10.58	17.54 nM	7.756
OA5	OAc	OH	H	H	CH_3	-12.01	1.57 nM	8.804
OA6	OMe	OH	H	H	CH_3	-12.21	1.13 nM	8.947
OA7	NH_2	OH	H	H	CH_3	-10	46.86 nM	7.329

OA8	Cl	OH	H	H	CH_3	-11.42	4.26 nM	8.37
OA9	OH	Cl	H	H	CH_3	-12.74	456.28 pM	9.341
OA10	OH	OMe	H	H	CH_3	-11.92	1.83 nM	8.737
OA11	OH	NH_2	H	H	CH_3	-11.03	8.27 nM	8.082
OA12	OH	OAc	H	H	CH_3	-12.28	988.85 pM	9.004
OA13	OH	H	H	OAc	CH_3	-11.27	5.50 nM	8.26
OA14	OH	H	H	NH_2	CH_3	-11.01	8.52 nM	8.069
OA15	OH	H	H	Cl	CH_3	-11.56	3.33 nM	8.477
OA16	OH	H	H	OMe	CH_3	-10.92	9.85 nM	8.006
OA17	OH	H	H	C_2H_5	CH_3	-11.07	7.67 nM	8.115
OA18	OH	H	H	OH	OEt	-13.41	147.10 pM	9.832
OA19	OH	H	H	OH	Cl	-11.6	3.17 nM	8.499
OA20	OH	H	H	OH	OAc	-13.45	137.91 pM	9.86
OA21	OH	H	H	OH	OMe	-12.65	535.69 pM	9.271
OA22	OH	OH	H	OH	OMe	-12.11	1.33 nM	8.876
OA23	OH	OH	H	H	NH_2	-10.63	16.03 nM	7.795
OA24	OH	OH	OH	H	OEt	-13.63	102.04 pM	9.991

OA25	OH	OH	H	OH	OEt	-13.72	87.53 pM	10.058
OA26	OH	OH	H	Cl	OEt	-13.76	81.45 pM	10.089
OA27	OH	H	Cl	OH	OH	-12.98	304.39 pM	9.516
OA28	OH	H	Cl	CH_3	OAc	-13.98	66.25 pM	10.178
OA29	OH	H	CH_3	CH_3	OAc	-14.56	21.41 pM	10.669
OA30	OH	H	CH_3	CH_3	Cl	-11.7	2.65 nM	8.576

3.8.1 2D QSAR

Various 2D molecular descriptors like Individual (like molecular weight, slogp, H-Acceptor count, Xlogp, H-donor count, rotatble bond count, smr etc.), chi, chiV, path count, path cluster, polar surface area were calculated. Alignment Inependent descriptors were calculated too. Invariable columns were removed, and Variable columns were used as the independent variables for the building the model. Most favourable test and training set were created using the Sphere Exclusion Method, which ensures that the points in both the sets are uniformly distributed w.r.t. chemical and biological space [34]. Seven molecules were selected as the test set and remaining twenty-two as the training set.

Four different models were generated using two Variable Separation Methods (Stepwise Variable Selection, Simulated Annealing) and two Regression methods (Multiple, Partial Least Square). The models were analyzed, and all the output data and graphs were saved.

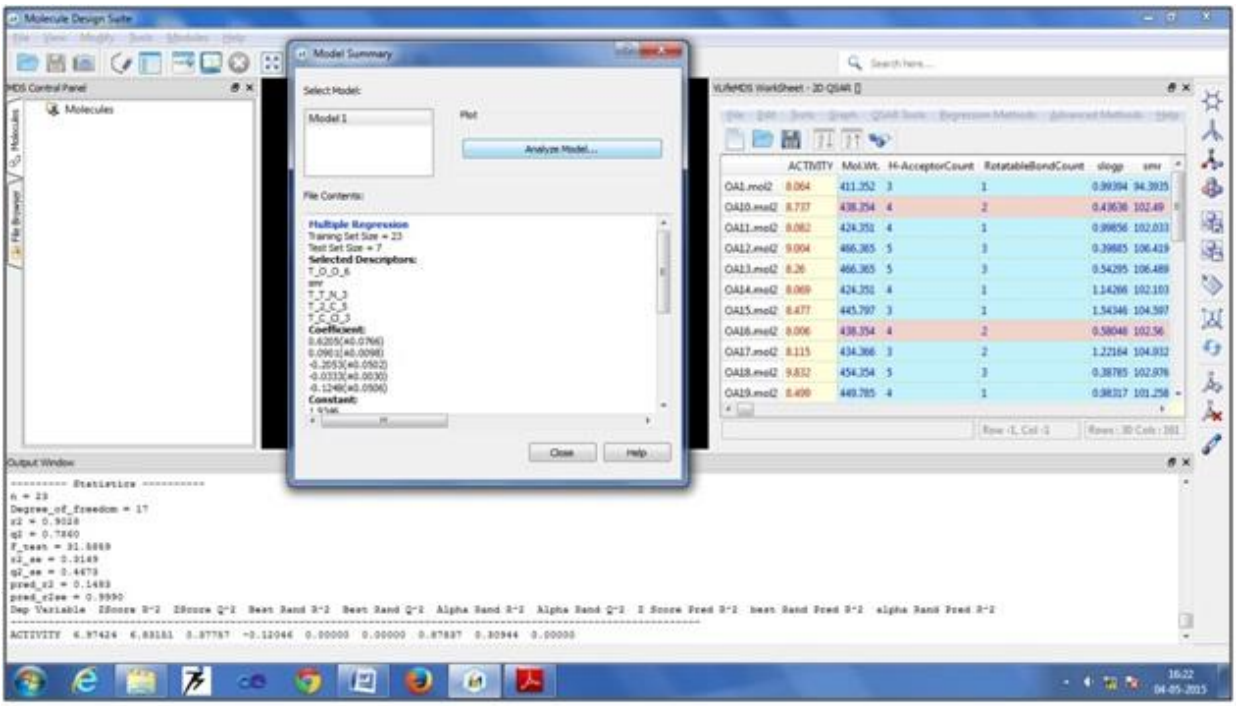

Figure 8: Model 1 - Multiple Regression with Forward-Backward Stepwise Variable Selection Method

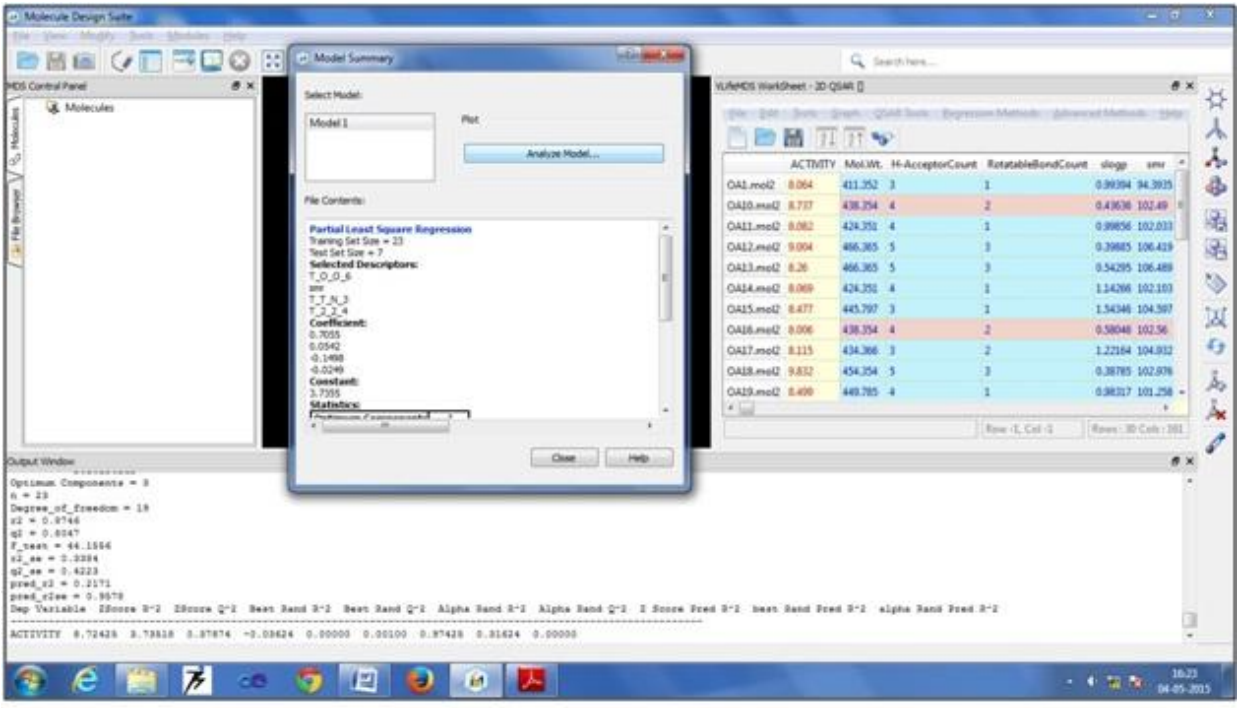

Figure 9: Model 2- Partial Least Square Regression with Forward- Backward Stepwise Variable Selection Method

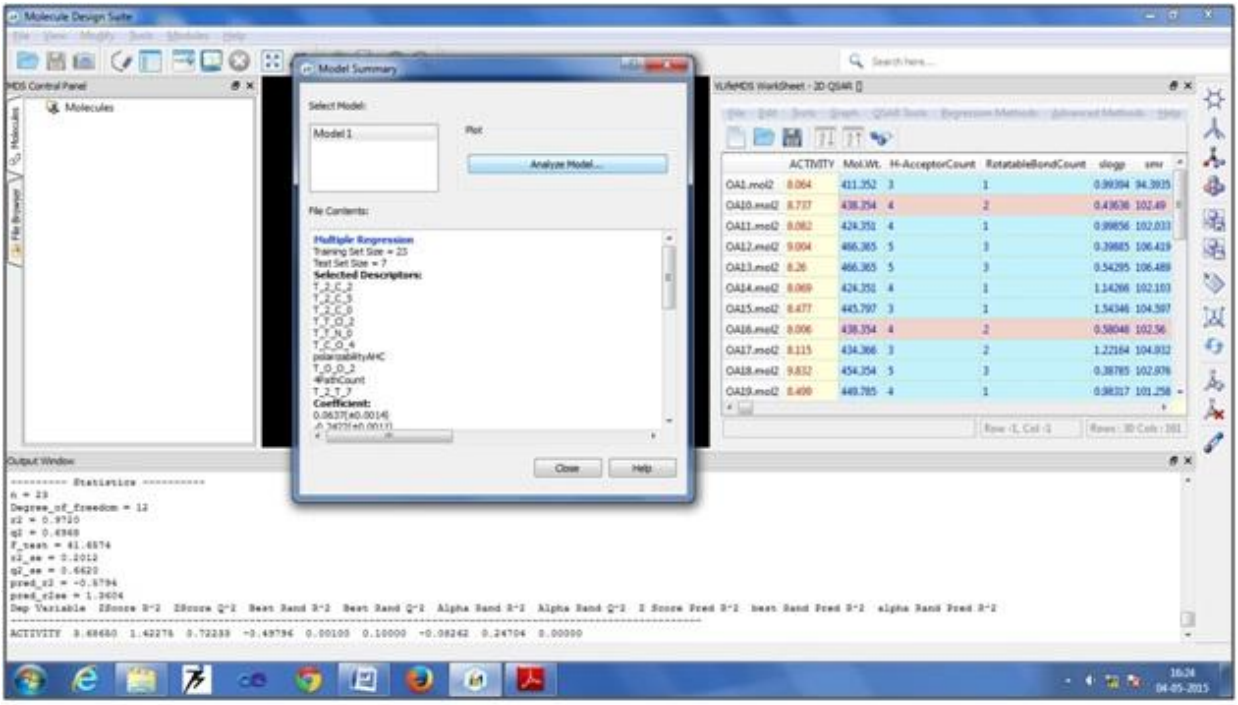

Figure 10: Model 3- Multiple Regression with Simulated Annealing Variable Selection Method

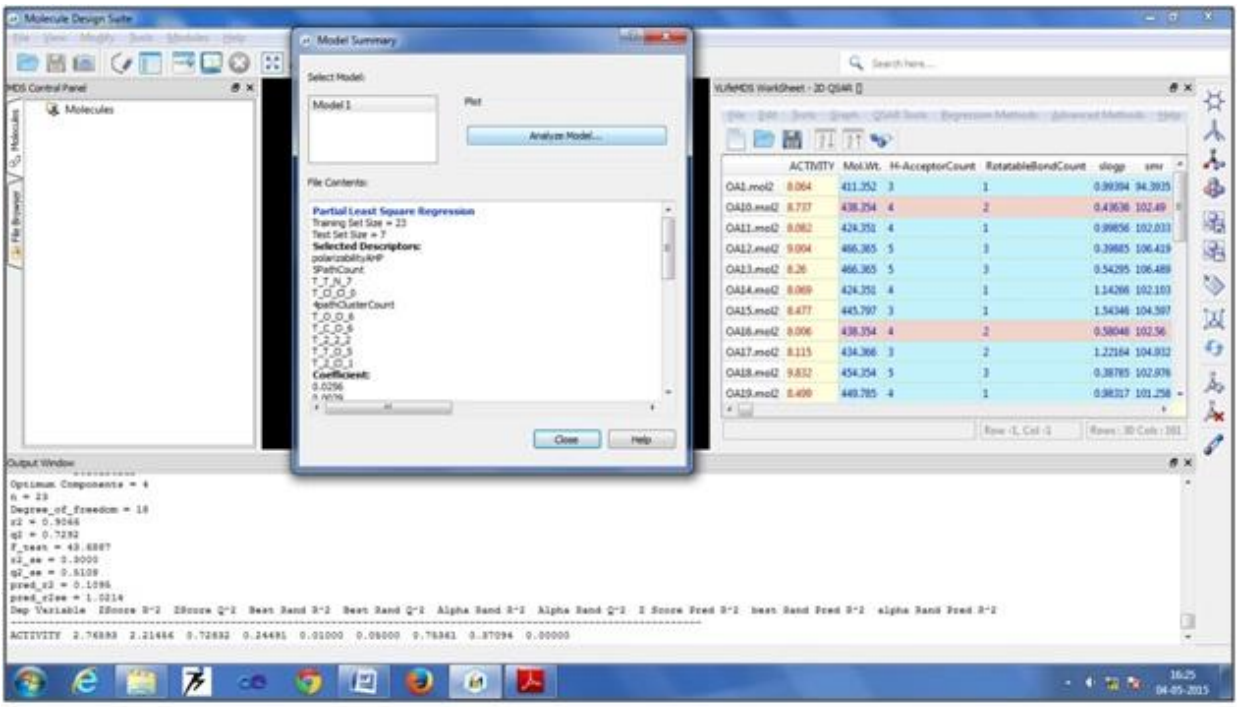

Figure 11: Model 4- Partial Least Square Regression with Simulated Annealing Variable Selection Method

3.8.2 ALIGNMENT OF MOLECULES

In a 2D QSAR model, the input data for building the model is independent for each molecule. However, in a 3D model the 3D descriptors like dipole moment, electrostatic and steric fields etc depend not only on individual molecular properties but the effect of each molecule on the others. Therefore, all the molecules in the data set for 3D QSAR need to be aligned with respect to each other. This helps in studying variations of each molecule with respect to the skeleton structure. Alignment of molecules was performed using the "Align Molecules" module of VLifeMDS version 4.3. Atom Selection Method was used, and the carbon atoms of the skeleton ring were set as the alignment atoms. The aligned molecules were saved and then opened in VLifeMDS by drag-n-drop to open simultaneously.

3.8.3 3D QSAR

3D QSAR was performed using the "3D QSAR" module of VLifeMDS software. 3D QSAR evaluates three-dimensional molecular fields and relates them to the activity of molecules. To build the 3D model, k-Nearest Neighbour method was used. In this method, a rectangular grid is generated around the set of aligned molecules. Molecular field energies are computed at the lattice points of the grid and are used as descriptors to establish the structure-activity relationship [35]. The electrostatic and steric fields were computed by applying the Gasteiger-Marsili charges and setting the dielectric constant as 1. The invariable columns were removed and the variable columns were used as the independent variable. Optimal test and training set were generated again using the Sphere Exclusion Method. 4 molecules were selected as the Data Set and the remaining 26 as the Training Set. The Unicolumn statistics like average, maximum, minimum etc of the activities of the Test and Training set were computed. The model was analyzed, and the output data and graphs were saved.

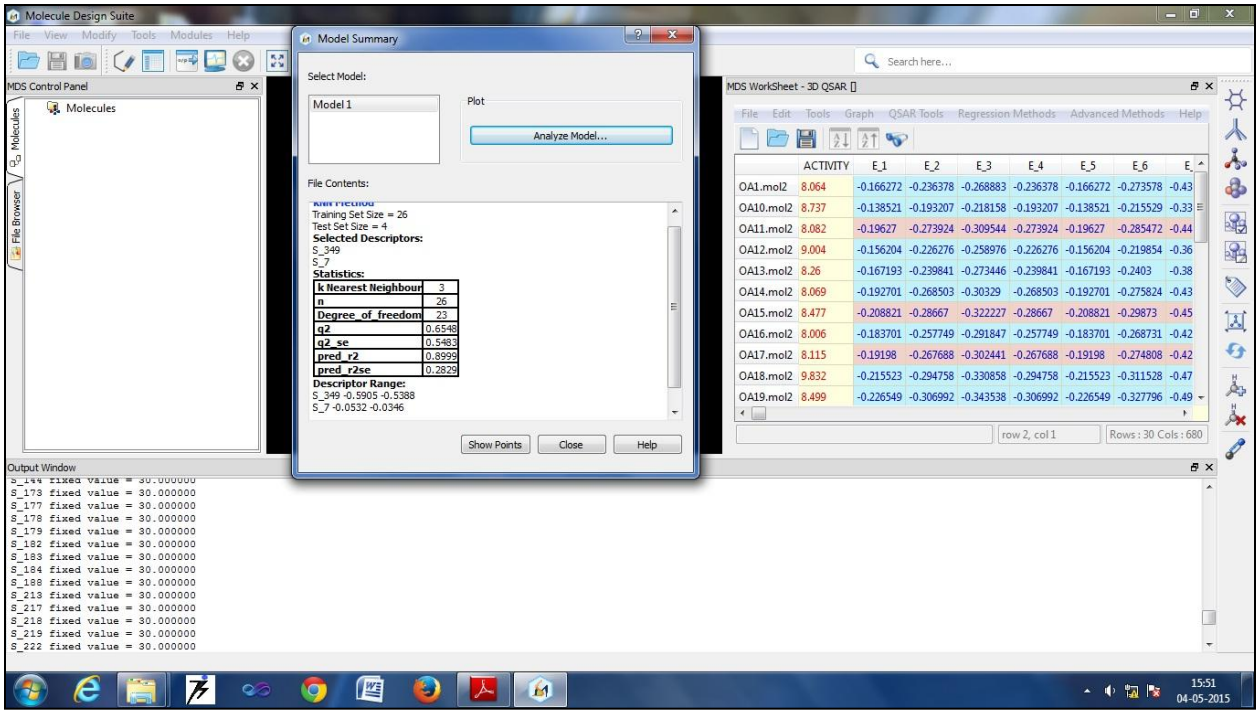

Figure 12: 3D Model – kNN method with Forward- Backward Stepwise Variable Selection Method

CHAPTER 4:

RESULTS AND DISCUSSIONS

4.1 RETRIEVAL AND PREPARATION OF THE RECEPTOR

The crystal structure of human TrkA receptor (high affinity NGFR) in complex with AZ-23 inhibitor (PDB ID 4AOJ), having resolution of 2.75 A° was retrieved from the Protein Data Bank and the miscellaneous ligands, the inhibitor and water molecules were removed using ArgusLab. Energy minimization and geometry optimization of the receptor were also done using ArgusLab.

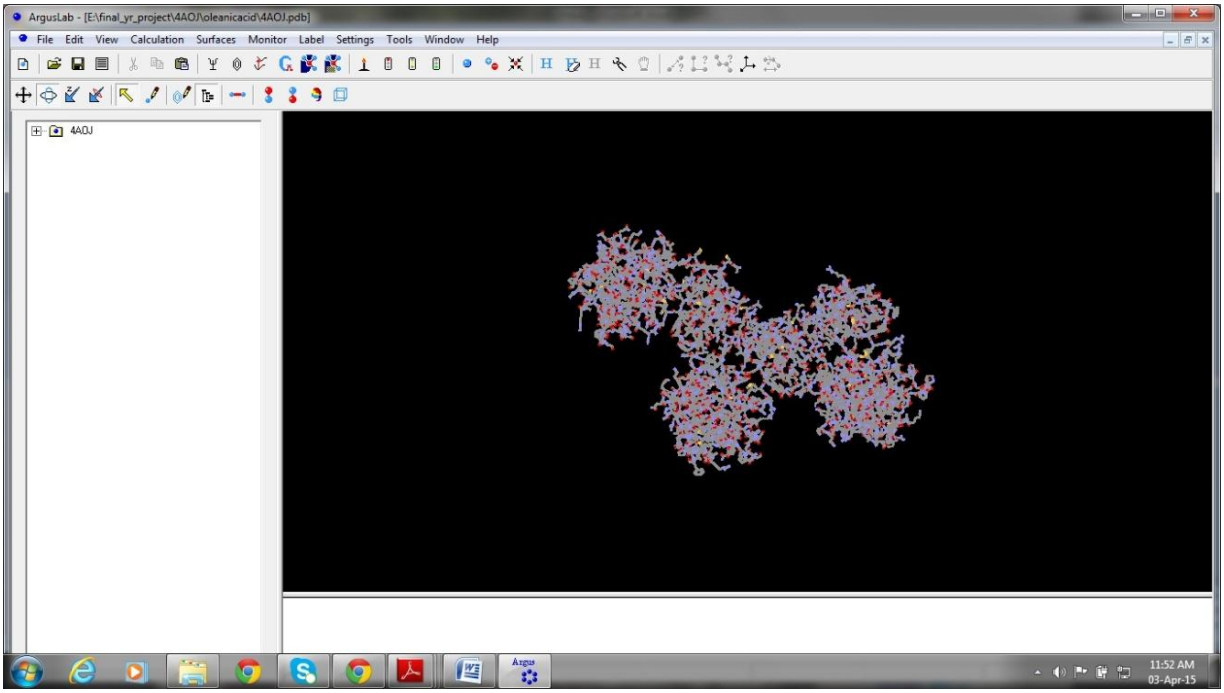

Figure 13: The receptor protein after required modifications in ArgusLab

4.2 DOCKING STUDIES

In our efforts to identify ligands modulating the NGFR, we performed molecular docking of a list of ligands with the target protein. Docking was performed using AutoDock 4.2 using the Lamarckian Genetic Algorithm. Lower the energy of binding, higher is the binding affinity of the ligand for the protein. Ligands with binding energy less than or equal to -8.00 kcal/mol are considered as the best compounds. We found that kaempferol-3-o-glucoside showed the highest binding affinity towards NGFR with a binding energy of -9.76 kcal/mol. Some other molecules like taraxerone, oleanolic acid, taraxerol, ursolic acid, alpha and beta-amyrin also showed significantly high binding affinity towards the target protein with

binding energies of -9.24 kcal/mol, -9.15 kcal/mol, -8.97 kcal/mol, -8.95 kcal/mol, -9.08 kcal/mol -8.95 kcal/mol respectively.

The docking result of ligands with the target protein are listed below :-

Table 3: Docking results of ligands with B.E < -7.5 kcal/mol

LIGANDS	**MIN. B.E. (kcal/mol)**	**Ki**
kaempferol-3-glucoside	-9.76	70.59 nM
Taraxerone	-9.24	169.24 nM
oleanolic acid	-9.15	197.51 Nm
alpha-amyrin	-9.08	221.30 nM
Taraxerol	-8.97	267.71 nM
ursolic acid	-8.95	277.44 nM
beta-amyrin	-8.95	274.05 nM
epigallocatechin gallate	-8.93	283.11 nM
Cycloeicosane	-8.82	344.41 nM
alpha sitosterol	-8.78	369.51 nM
kaempferol-3-rutinoside	-8.76	379.76 nM
Hesperidin	-8.62	476.50 nM
Diosmetin	-8.6	497.36 nM
Luteolin	-8.56	534.72 nM
Stigmasterol	-8.52	569.50 nM
Quercetin	-8.5	585.01 nM
Eriodictyol	-8.46	628.13 nM
catechin gallate	-8.4	700.69 nM
Baicalin	-8.39	704.76 nM
Jatamansin	-8.37	731.16 Nm
gamma sitosterol	-8.33	784.89 nM
rosmariquinone (miltirone)	-8.23	931.41 nM

Reserpine	-8.2	972.55 nM
beta sitosterol	-8.19	984.40 nM
Dihydroquercetin	-8.18	1.02 μM
Kaempferol	-8.14	1.08 μM
Rutin	-8.12	1.11 μM
carnosic acid	-8.1	1.15 μM
Ajmalicine	-8.08	1.19 μM
Wogonin	-8.07	1.22 μM
Tabersonine	-8.03	1.30 μM
ginkgolide A	-8.01	1.34 μM
Hispidulin	-8.01	1.34 μM
Catharanthine	-7.95	1.49 μM
Carnosol	-7.94	1.51 μM
Baicalein	-7.92	1.57 μM
Nomilin	-7.9	1.61 μM
Galangin	-7.86	1.72 μM
Myrcetin	-7.82	1.87 μM
quercetin 3-methyl ether	-7.8	1.93 μM
valerenic acid	-7.74	2.13 μM
ginkgolide B	-7.74	2.11 μM
chlorogenic acid	-7.71	2.22 μM
Apigenin	-7.6	2.69 μM
Taxifolin	-7.57	2.82 μM
ginkgolide J	-7.57	2.84 μM
Norwogonin	-7.56	2.90 μM
Oroxylin	-7.55	2.92 μM
Harmaline	-7.53	3.05 μM
Hydroxypinoresinol	-7.53	3.02 μM

4.3 ADME RESULTS

The ADME properties predictions were done using the online PreADMET server. The PreADMET server calculates the CACO2 (human colon adenocarcinoma) model, MDCK (Madin-Darby Canine Kidney) model, Human Intestinal Absorption (HIA) which provide a measure of absorption; and Skin Permeability and Blood-Brain Barrier (BBB) models which give a measure of distribution in the body system.

Table 4: Optimum values of Absorption and Distribution Models

	Low Absorption	**Moderate**	**High**
CACO2	<4 nm/sec	4-70 nm/sec	>70 nm/sec
MDCK	<25 nm/sec	25-500 nm/sec	>500 nm/sec
HIA	<20 %	20-70 %	>70 %
	Low Distribution	**Moderate**	**High**
BBB (C_{brain}/C_{blood})	<0.1	0.1-2	>2
Plasma-Protein Binding	>90 %		<90 %

Table 5: ADME Properties of ligands having high binding affinity to the target protein

LIGAND	**CACO2 (nm/sec)**	**MDCK (nm/sec)**	**HIA (%)**	**SKIN PERMEABI-LITY (log K_p)**	**BBB**	**PLASMA-PROTEIN BINDING (%)**
Kaempferol-3-glucoside	11.1458	1.1475	25.1716	-4.6476	0.03500	57.5756
Taraxerone	49.5959	0.1641	100	-1.9651	21.2078	100
oleanic acid	21.8872	0.0439	95.9963	-2.3543	7.8792	100
alpha-amyrin	49.0784	1.0017	100	-1.8331	18.3975	100
Taraxerol	48.7377	0.1417	100	-1.8570	18.5142	100
ursolic acid	22.0031	0.0449	96.0292	-1.5881	7.5231	100
beta-amyrin	46.7500	0.1749	100	-2.2225	21.2500	100

epigallocatechin gallate	12.0421	0.0446	20.7125	-3.9954	0.0875	100
Cycloeicosane	22.1927	68.0213	100	-0.5025	20.4647	100
alpha-sitosterol	52.6295	4.1754	100	-0.5928	19.9303	100
k3rutinoside	9.1328	0.2510	6.2896	-4.5985	0.0293	42.3746
Hesperidin	5.9216	0.0815	9.8828	-4.6202	0.0300	43.4520
Diosmetin	7.0252	23.8531	88.1882	-4.1347	0.2011	90.1601
Luteolin	4.5397	36.5205	79.4272	-4.2801	0.3676	99.7172
Stigmasterol	52.3734	8.8572	100	-0.5934	19.8883	100
Quercetin	3.4129	13.3528	63.4852	-4.4334	0.1727	93.2361
Eriodictyol	4.5336	35.5683	77.4301	-4.3140	0.3802	100
catechingallate	13.2124	0.0460	40.5818	-3.8897	0.1411	100
Baicalin	11.5594	0.1477	32.4230	-4.3741	0.0252	75.6919
Jatamansin	39.4601	3.4499	98.5170	-2.6101	0.1507	93.8556
Gsitosterol	52.3734	8.8572	100	-0.5934	19.8883	100
Miltirone	23.9684	89.6944	93.7125	-1.6150	10.0165	100
Reserpine	35.0577	0.0436	89.7527	-4.7246	0.2252	54.3845
Bsitosterol	52.3734	8.8572	100	-0.5934	19.8883	100
dihydroquercetin	3.4231	9.5674	60.1637	-4.4261	0.1669	95.1644
Kaempferol	9.5774	29.6119	79.4393	-4.3255	0.2861	89.6082
Rutin	7.91267	0.3269	2.8612	-4.6667	0.0285	43.8979
carnosic acid	20.7221	120.711	92.5232	-1.5304	4.1091	96.6142
Ajmalicine	20.7542	5.1815	90.6079	-4.8837	1.4696	29.5477
Wogonin	4.2822	152.119	93.0394	-3.3113	0.7246	90.4470
Tabersonine	35.1299	6.3901	91.5681	-4.5715	2.69073	46.3189
Ginkgolidea	20.6965	0.8771	64.7994	-4.8812	0.13814	49.2247
Hispidulin	5.4632	32.2235	88.1840	-4.1299	0.1073	91.5559
catharanthine	34.6456	37.9216	91.8445	-4.5271	5.2094	55.8748
Carnosol	21.1724	144.317	92.5807	-2.6397	4.5820	97.5762

Baicalein	1.2802	101.909	88.1054	-4.1355	0.7708	98.9832
Nomilin	25.4412	0.6416	94.4843	-3.3555	0.1102	66.6485
Galangin	3.7059	79.4051	88.1226	-4.1908	0.6713	88.1221
Myrcetin	0.9913	0.9913	40.9640	-4.5272	0.1103	96.7848
Kaempferol-3-methylether	3.9330	32.8102	88.1926	-4.2555	0.3009	83.9814
valerenic acid	15.1311	154.204	98.3301	-1.3518	0.9793	100
Ginkgolideb	20.572	0.6127	40.3047	-5.0305	0.1193	41.2074
chlorogenic acid	18.7316	4.267	17.5366	-3.9402	0.0372	41.4260
Apigenin	10.5468	44.302	88.1228	-4.1457	0.5651	97.2534
Taxifolin	3.42307	9.5674	60.1637	-4.4261	0.1669	95.1644
Ginkgolidej	20.8757	0.6279	40.3047	-5.0421	0.1143	41.4346
Norwogonin	2.0536	108.04	88.1095	-4.1254	0.8388	96.6339
Oroxylin	1.7836	129.427	93.0362	-3.3254	0.1235	91.2836
Harmaline	40.4085	307.684	90.9619	-4.4240	4.9845	72.7077
Hydroxypinoresinol	22.792	4.3395	89.6144	-4.1818	0.0942	69.8990

The above table reveals that though kaempferol-3-glucoside is the best docked ligand, its ADME properties are relatively poor. Kaempferol-3-glucoside has low values of CACO2 and MDCK models indicating low oral absorption. It has a low value for Human intestinal absorption as well and low value of BBB indicates that it has low distribution to the Central Nervous System. However, the next set of molecules – taraxerone, oleanolic acid, taraxerol, alpha and beta-amyrin and ursolic acid were found to have high values for CACO2 and MDCK models. They even have high values of human intestinal absorption and BBB indicating high distribution to the Central Nervous System. Thus, these molecules could be lead compounds for targeting the NGFR.

4.4 QSAR STUDIES

Biological activity of any compound is directly linked to its molecular structure. Hence, compounds with similar structure are expected to possess similar bioactivities. QSAR is a computational method to build a mathematical model that can help to predict the bioactivities so similar compounds. The mathematical model establishes a correlation between various molecular properties (descriptors) of a set of molecules with their experimentally known biological activity. 2D and 3D QSAR were performed with 30 oleanolic acid derivatives and various models were generated.

4.4.1 2D QSAR

Certain 2D descriptors like - Individual (like molecular weight, slogp, H-Acceptor count, Xlogp, H-donor count, rotatble bond count, smr etc.), chi, chiV, path count, path cluster, polar surface area were calculated.

Table 6: Some 2D molecular descriptor values

	Activity	**Mol.wt.**	**H-Acceptor count**	**Rotatable Bond count**	**Slogp**	**Smr**	**Polarizability AHC**
OA1.mol2	8.064	411.352	3	1	0.994	94.394	10.435
OA10.mol2	8.737	438.354	4	2	0.436	102.49	17.992
OA11.mol2	8.082	424.351	4	1	0.999	102.033	17.137
OA12.mol2	9.004	466.365	5	3	0.399	106.42	21.493
OA13.mol2	8.26	466.365	5	3	0.543	106.49	21.493
OA14.mol2	8.069	424.351	4	1	1.143	102.103	17.137
OA15.mol2	8.477	445.797	3	1	1.543	104.597	19.978
OA16.mol2	8.006	438.354	4	2	0.58	102.56	17.992
OA17.mol2	8.115	434.366	3	2	1.222	104.932	18.323
OA18.mol2	9.832	454.354	5	3	0.388	102.977	18.796
OA19.mol2	8.499	449.785	4	1	0.983	101.259	19.66
OA2.mol2	8.368	426.343	4	1	0.661	99.16	16.924

OA20.mol2	9.86	470.353	6	3	-0.017	103.151	21.169
OA21.mol2	9.271	442.343	5	2	0.02	99.221	17.696
OA22.mol2	8.876	457.334	6	2	-0.464	97.489	16.687
OA23.mol2	7.795	428.339	5	1	0.619	97.375	16.698
OA24.mol2	9.991	469.345	6	3	0.475	100.626	17.738
OA25.mol2	10.058	469.345	6	3	-0.096	101.245	17.759
OA26.mol2	10.089	488.799	5	3	1.146	107.463	19.209
OA27.mol2	9.516	488.799	5	3	0.605	108.001	20.758
OA28.mol2	10.178	500.81	5	3	1.039	111.486	21.854
OA29.mol2	10.669	477.368	5	3	0.636	108.578	19.112
OA3.mol2	10.55	466.365	5	3	0.576	104.63	18.27
OA30.mol2	8.576	456.8	3	1	1.637	106.685	17.637
OA4.mol2	7.756	440.327	5	1	0.054	96.371	15.658
OA5.mol2	8.804	466.365	5	3	0.866	105.612	19.953
OA6.mol2	8.947	438.354	4	2	0.904	101.683	16.542
OA7.mol2	7.329	424.351	4	1	1.106	100.444	17.055
OA8.mol2	8.37	445.797	3	1	1.867	103.72	18.459
OA9.mol2	9.341	445.797	3	1	1.399	104.527	19.978

The invariable columns do not have any contribution to the model building and hence were removed. The Unicolumn statistics of the remaining columns were found to be as tabulated below.

Table 7: Unicolumn Statistics

Column name	**Average**	**Maximum**	**Minimum**	**StdDev**	**Sum**
ACTIVITY (Training Set)	8.7857	10.6690	7.3290	0.8879	202.0720
ACTIVITY (Test Set)	9.3453	10.5500	8.0060	0.8980	65.4170

The maximum of Test Set was found to be less than the maximum of Training Set and minimum of Test Set was found to be less than the minimum of Training Set, which is in accordance with the set guidelines [36].

Four 2D models were generated by taking different Variable Selection Methods and Regression Methods.

Table 7: Model equations of the four 2D models

MODEL	MODEL EQUATION
Model 1	ACTIVITY = + 0.6205(± 0.0766) T_O_O_6 + 0.0901(± 0.0098) smr - 0.2053(± 0.0502) T_T_N_3 - 0.0333(± 0.0030) T_2_C_5 - 0.1248(± 0.0506) T_C_Cl_3 + 1.9346
Model 2	ACTIVITY = + 0.7055 T_O_O_6 + 0.0542 smr - 0.1498 T_T_N_3 - 0.0249 T_2_2_4 + 3.7355
Model 3	ACTIVITY = + 0.0637(± 0.0014) T_2_C_2 - 0.2422(± 0.0011) T_2_C_5 + 0.2246(± 0.0219) T_2_C_0 - 0.4624(± 0.0252) T_T_Cl_2 - 1.7685(± 0.1287) T_T_N_0 - 0.1654(± 0.0034) T_C_O_4 + 0.0286(± 0.0046) polarizabilityAHC - 0.3945(± 0.1112) T_O_O_2 + 0.0234(± 0.0020) 4PathCount + 0.1696(± 0.0007) T_2_T_7 + 7.8875
Model 4	ACTIVITY = + 0.0256 polarizabilityAHP + 0.0029 5PathCount - 0.3047 T_T_N_7 - 0.0052 T_Cl_Cl_0 + 0.0879 4pathClusterCount + 0.6786 T_O_O_6 - 0.0692 T_C_O_6 - 0.0454 T_2_2_2 + 0.0160 T_T_O_5 - 0.2454 T_2_Cl_1 + 5.1757

MODEL 1:

ACTIVITY = + 0.6205(± 0.0766) T_O_O_6 + 0.0901(± 0.0098) smr - 0.2053(± 0.0502) T_T_N_3 - 0.0333(± 0.0030) T_2_C_5 - 0.1248(± 0.0506) T_C_Cl_3 + 1.9346

N_{train} = 23, N_{test} = 7, Degree of freedom = 17, r^2 = 0.9028, q^2 = 0.7860, F_test = 31.5859, r^2_se = 0.3149, q^2_se = 0.4673, pred_r^2 = 0.1483, pred_r^2se = 0.9990

The Model 1, with Multiple Regression Method and Forward-Backward Stepwise Variable Selection Method, has the coefficient of determination (r^2) value of 0.9028. It shows cross-validated squared correlation coefficient (q^2 = 0.7860) of 79% and a predictivity for the external test set (pred_r^2 = 0.1483) of 15%. The coefficient magnitude of a descriptor shows its relative contribution with respect to other descriptors and sign indicates whether it is directly or inversely proportional to the activity. In this QSAR model, the positive coefficient of T_O_O_6 and smr shows that an increase in the value of these descriptors enhances the activity. On the other hand, negative coefficients of T_T_N_3, T_2_C_5 and T_C_Cl_3 indicate that an increase in these descriptor values will decrease the activity of the compound.

The descriptor T_O_O_6, which is the count of number of Oxygen atoms (single, double or triple bonded) separated from another Oxygen atom by 6 bonds, plays the most important role in the activity of the compound with a contribution of approximately 34%. The descriptor smr, which signifies the molecular refractivity which in turn is a measure of molecular size, contributes approximately 22% in determining the activity. T_T_N_3 (count of any atom separated from Nitrogen atom by 3 bonds), T_2_C_5 (count of any type of double bonded atom separated from Carbon atom by 5 bonds) and T_C_Cl_3 (count of Carbon atom separated from Chlorine atom by 3 bonds) have negative contributions of approximately 18%, 15% and 12% respectively.

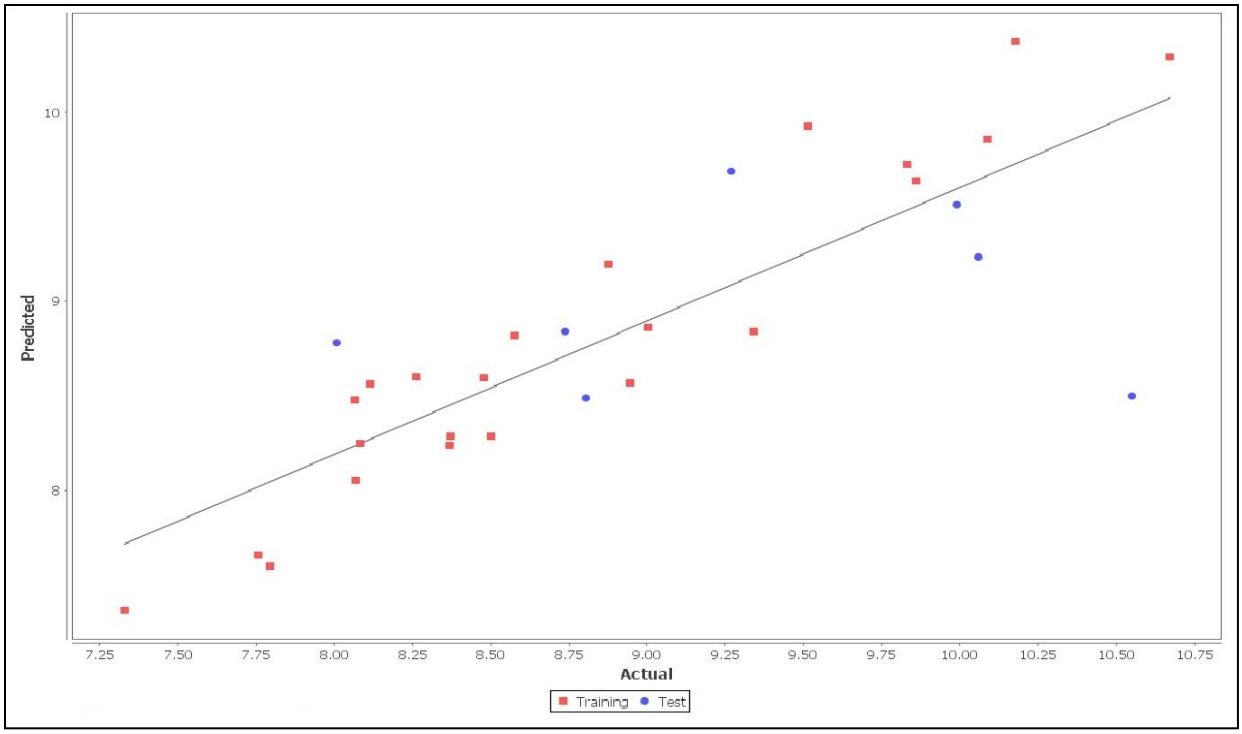

Figure 14: Fitness Plot of 2D QSAR Model 1

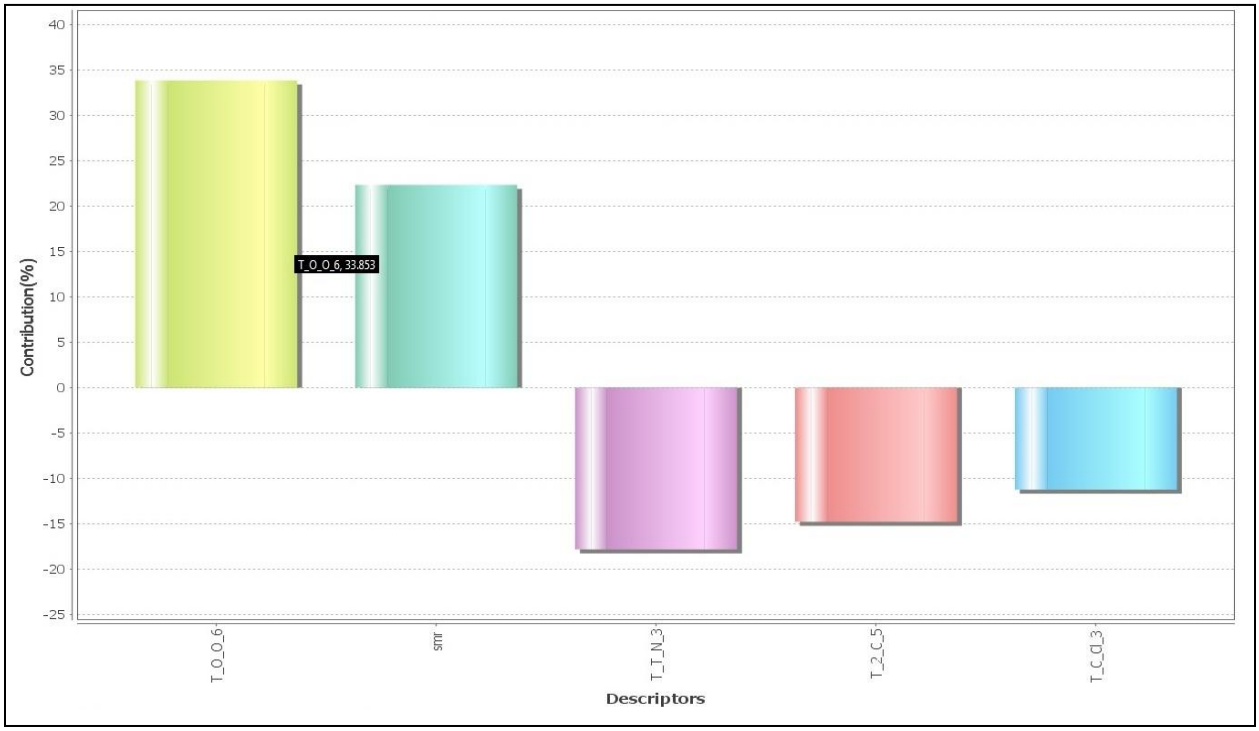

Figure 15: Contribution Plot 2D QSAR Model 1

MODEL 2:

ACTIVITY = + 0.7055 T_O_O_6 + 0.0542 smr - 0.1498 T_T_N_3 - 0.0249 T_2_2_4 + 3.7355

N_{train} = 23, N_{test} = 7, Degree of freedom = 19, r^2 = 0.8746, q^2 = 0.8047, F_test = 44.1556, r^2_se = 0.3384, q^2_se = 0.4223, pred_r^2 = 0.2171, pred_r^2se = 0.9578

In model 2 (with Partial Least Square Regression and Forward-Backward Stepwise Variable Selection), the coefficient of determination (r^2) value was computed to be 0.8746. It shows cross-validated squared correlation coefficient (q^2 = 0.8047) of 80% and a predictivity for the external test set (pred_r^2 = 0.2171) of 22%. The positive coefficient of T_O_O_6 and smr indicate that an increase in these descriptors increases the activity and the negative coefficients of T_T_N_3 and T_2_2_4 imply that an increase in these descriptors will decrease the activity. According to this 2D model, the descriptor T_O_O_6, the count of number of Oxygen atoms (single, double or triple bonded) separated from another Oxygen atom by 6 bonds, contributes to the activity by approximately 51% . The descriptor smr, which is the molecular refractivity which (a measure of molecular size), contributes approximately 17% in determining the activity. T_T_N_3(any atom separated from a Nitrogen atom by 3 bonds) and T_2_2_4 (any double bonded atom separated from another double bonded atom by 4 bonds) descriptors have negative contributions of 17% and 14% respectively.

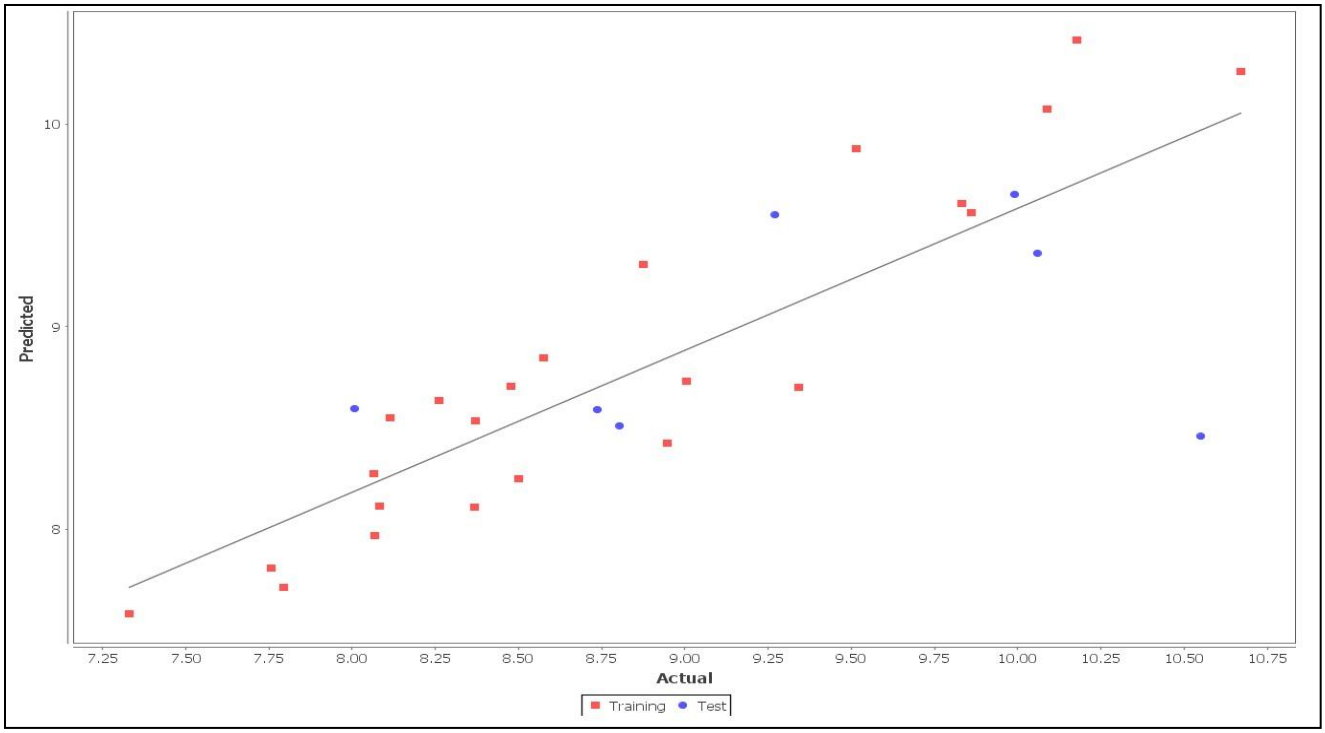

Figure 16: Fitness Plot of 2D QSAR Model 2

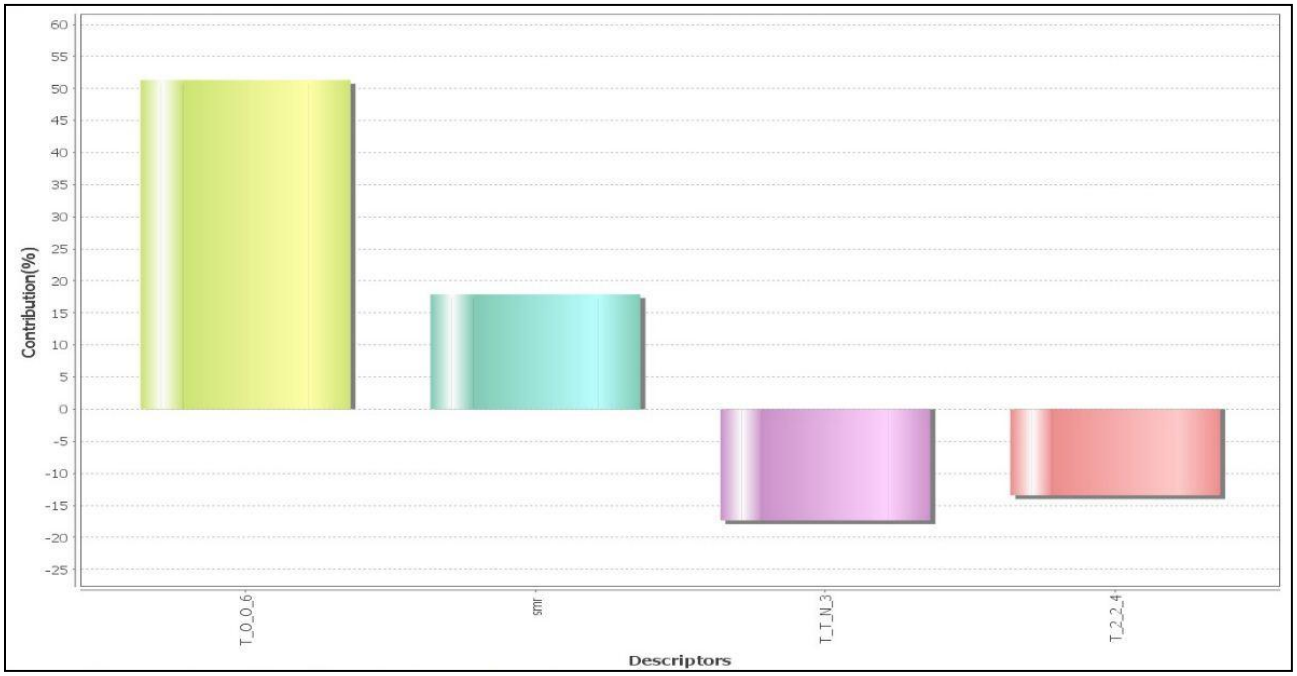

Figure 17: Contribution Plot of 2D QSAR Model 2

MODEL 3:

ACTIVITY = + 0.0637(± 0.0014) T_2_C_2 - 0.2422(± 0.0011) T_2_C_5 + 0.2246(± 0.0219) T_2_C_0 - 0.4624(± 0.0252) T_T_Cl_2 - 1.7685(± 0.1287) T_T_N_0 - 0.1654(± 0.0034) T_C_O_4 + 0.0286(± 0.0046) polarizabilityAHC - 0.3945(± 0.1112) T_O_O_2 + 0.0234(± 0.0020) 4PathCount + 0.1696(± 0.0007) T_2_T_7 + 7.8875

N_{train} = 23, N_{test} = 7, Degree_of_freedom = 12, r^2 = 0.9720, q^2 = 0.6968, F_test = 41.6574, r2_se = 0.2012, q2_se = 0.6620, pred_r2 = 0.5794, pred_r2se = 1.3604

The Model 3, with Multiple Regression Method and Simulated Annealing Variable Selection Method, has the coefficient of determination (r^2) value of 0.9720. It shows cross-validated squared correlation coefficient (q^2 = 0.6968) of 70% and a predictivity for the external test set (pred_r^2 = 0.5794) of 58%. In this QSAR model, positive sign of T_2_C_2, T_2_C_0, polarizabilityAHC, 4PathCount and T_2_T_7 indicate that an increase in these descriptors increases the activity of the compound. On the other hand, negative sign of T_2_C_5, T_T_Cl_2, T_T_N_0, T_C_O_4 and T_O_O_2 signify that an increase in these descriptors decrease the activity of the compound. The descriptor T_2_C_2, the count of any double bonded atoms separated from Carbon atom by 2 bonds, contributes to the activity by 6%. The descriptor T_2_C_0, the count of any atom double bonded to Carbon atom, contributes to the activity by 5%. The descriptor polarizabilityAHC, which is a measure of molecular polarizability has a contribution of approximately 1%. The descriptor 4PathCount, which signifies the number of fragments of fourth order, has a contribution of approximately 2%. The descriptor T_2_T_7, which is the count of double bonded atoms separated from any atom by 7 bonds, has a contribution of approximately 26%. The descriptors T_2_C_5 (count of double bonded atoms separated from Carbon atom by 5 bonds), T_T_Cl_2 (count of any atom separated from Chlorine atom by 2 bonds), T_T_N_0 (count of any atom bonded to Nitrogen atom), T_C_O_4 (count of Carbon atoms separated from Oxygen atoms by 4 bonds) and T_O_O_2 (count of oxygen atoms separated from oxygen atoms by 2 bonds) have negative contributions of approximately 27%, 7%, 10%, 9% and 3% respectively.

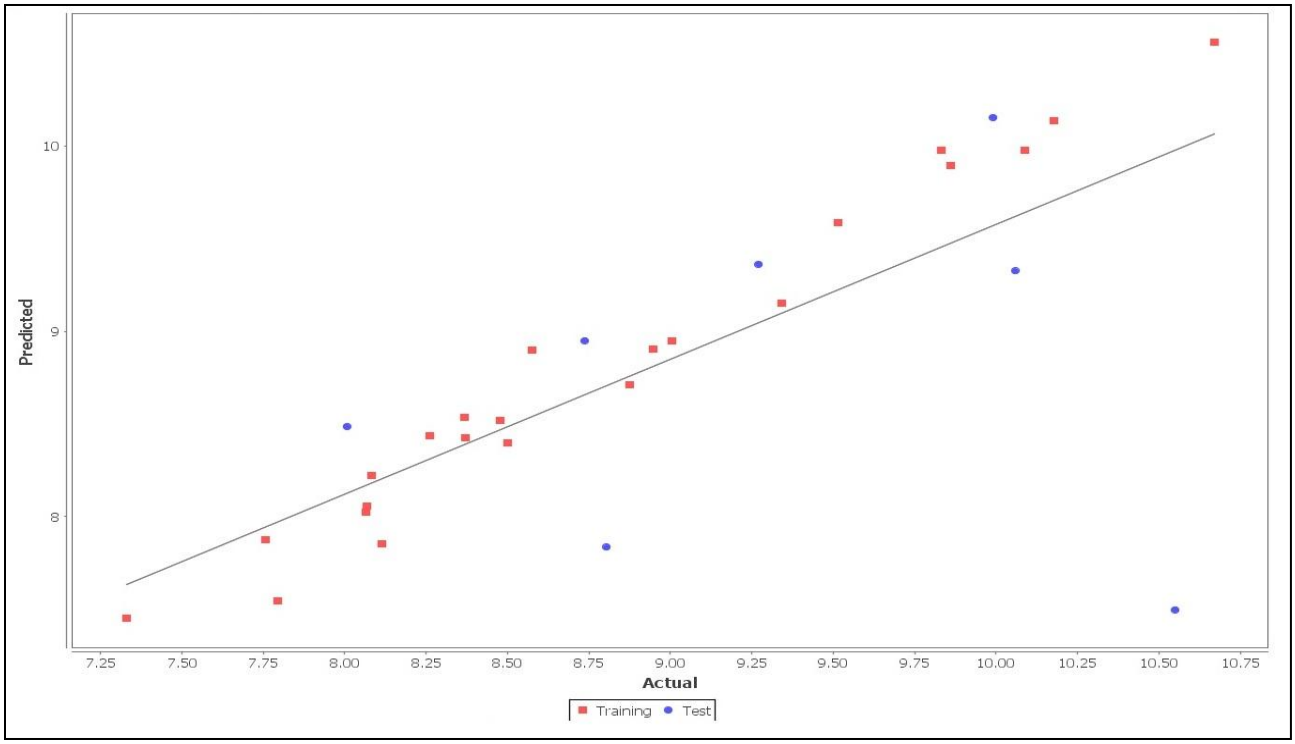

Figure 18: Fitness Plot of 2D QSAR Model 3

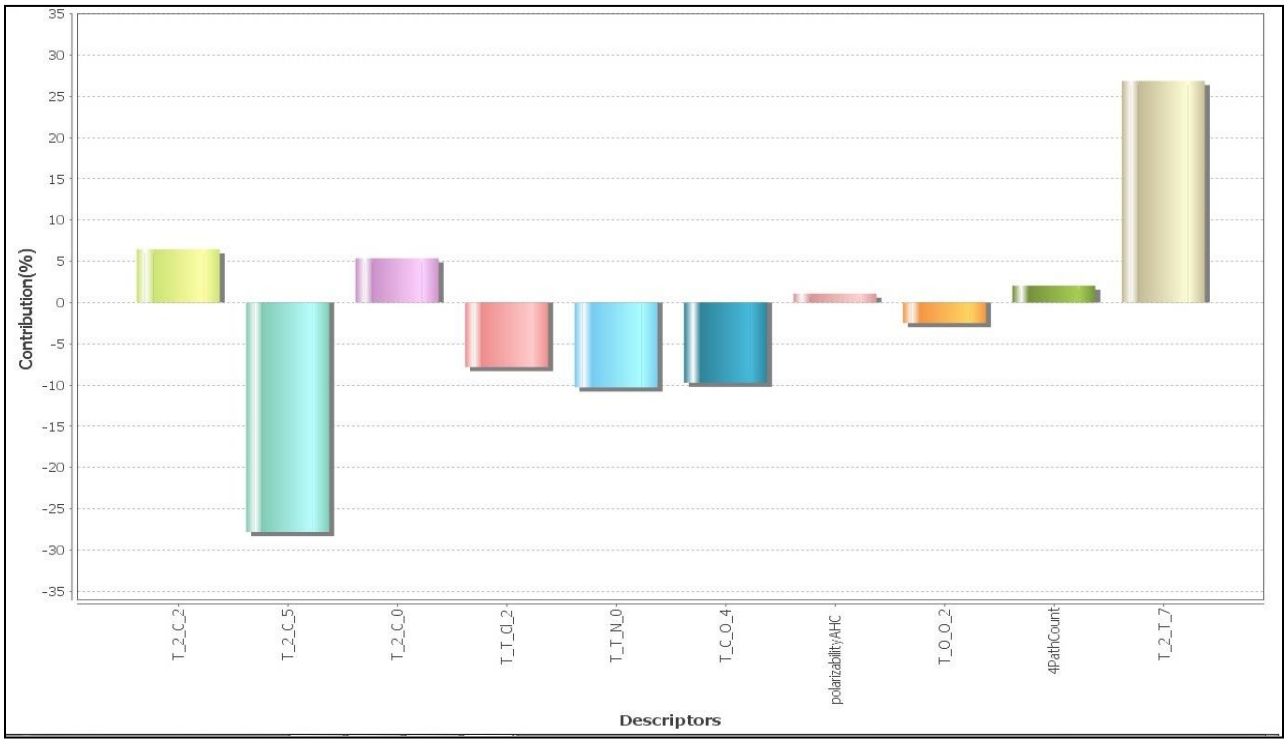

Figure 17: Contribution Plot of 2D QSAR Model3

MODEL 4:

ACTIVITY = + 0.0256 polarizabilityAHP + 0.0029 5PathCount - 0.3047 T_T_N_7 - 0.0052 T_Cl_Cl_0 + 0.0879 4pathClusterCount + 0.6786 T_O_O_6 - 0.0692 T_C_O_6 - 0.0454 T_2_2_2 + 0.0160 T_T_O_5 - 0.2454 T_2_Cl_1 + 5.1757

N_{train} = 23, N_{test} = 7, Degree_of_freedom = 18, r^2 = 0.9066, q^2 = 0.7292, F_test = 43.6887, r2_se = 0.3000, q2_se = 0.5108, pred_r2 = 0.1095, pred_r2se = 1.0214

The Model 4, with Partial Least Square Regression Method and Simulated Annealing Variable Selection Method, has the coefficient of determination (r^2) value of 0.9066. It shows cross-validated squared correlation coefficient (q^2 = 0.7292) of 73% and a predictivity for the external test set (pred_r^2 = 0.1095) of 11%. In this QSAR model, positive sign of polarizabilityAHP, 5PathCount, 4pathClusterCount, T_O_O_6 and T_T_O_5 signifies that an increase in these descriptors increases the activity of the compound. On the other hand, negative sign of T_T_N_7, T_Cl_Cl_0, T_O_O_6, T_C_O_6, T_2_2_2 and T_2_Cl_1 indicates that an increase in these descriptors decreases the activity of the compound. The descriptor polarizabilityAHP, which is a measure of molecular polarizability has a contribution of approximately 3%. The descriptor 5PathCount, which signifies the number of fragments of fifth order has a contribution of approximately 1%. The descriptor 4pathClusterCount, which signifies the number of fragments of fourth order pathcluster in a molecule, has a contribution of approximately 15%. The descriptor T_O_O_6, which is the count of oxygen atoms separated from other oxygen atoms by 6 bonds, has the highest contribution of approximately 32%. The descriptor T_T_O_5, which is the count of atoms separated from Oxygen atoms by 5 bonds, contributes by approximately 5%. The descriptors T_T_N_7 (count of atoms separated from Nitrogen atoms by 7 bonds), T_Cl_Cl_0 (count of Chlorine atoms bonded to other Chlorine atoms), T_O_O_6 (count of Oxygen atoms separated from other oxygen atoms by 6 bonds), T_C_O_6 (count of Carbon atoms separated from Oxygen atoms by 2 bonds), T_2_2_2 (count of any double bonded atom separated by 2 bonds from another double bonded atom) and T_2_Cl_1 (count of any double bonded atom separated from Chlorine atom by 1 bond) have negative contributions of approximately 17%, 1%, 11%, 11% and 4% respectively.

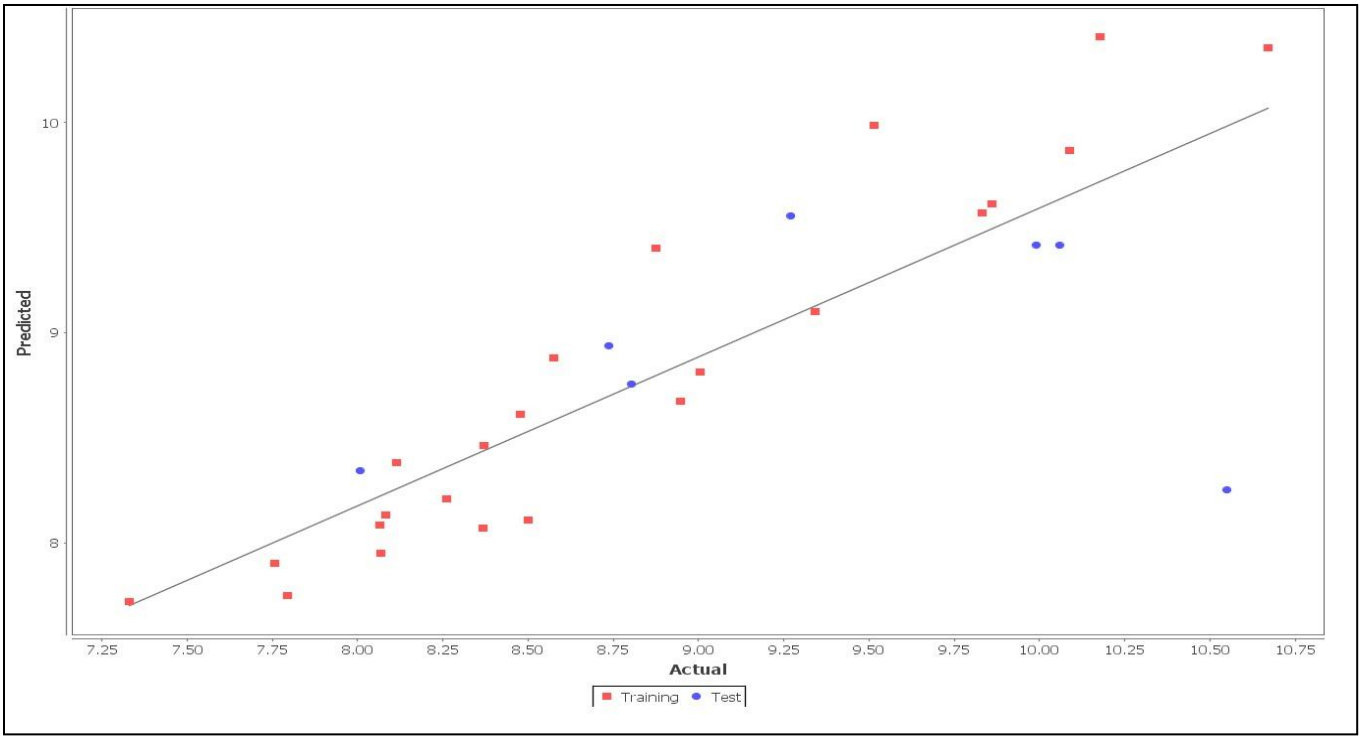

Figure 20: Fitness Plot of 2D QSAR Model 4

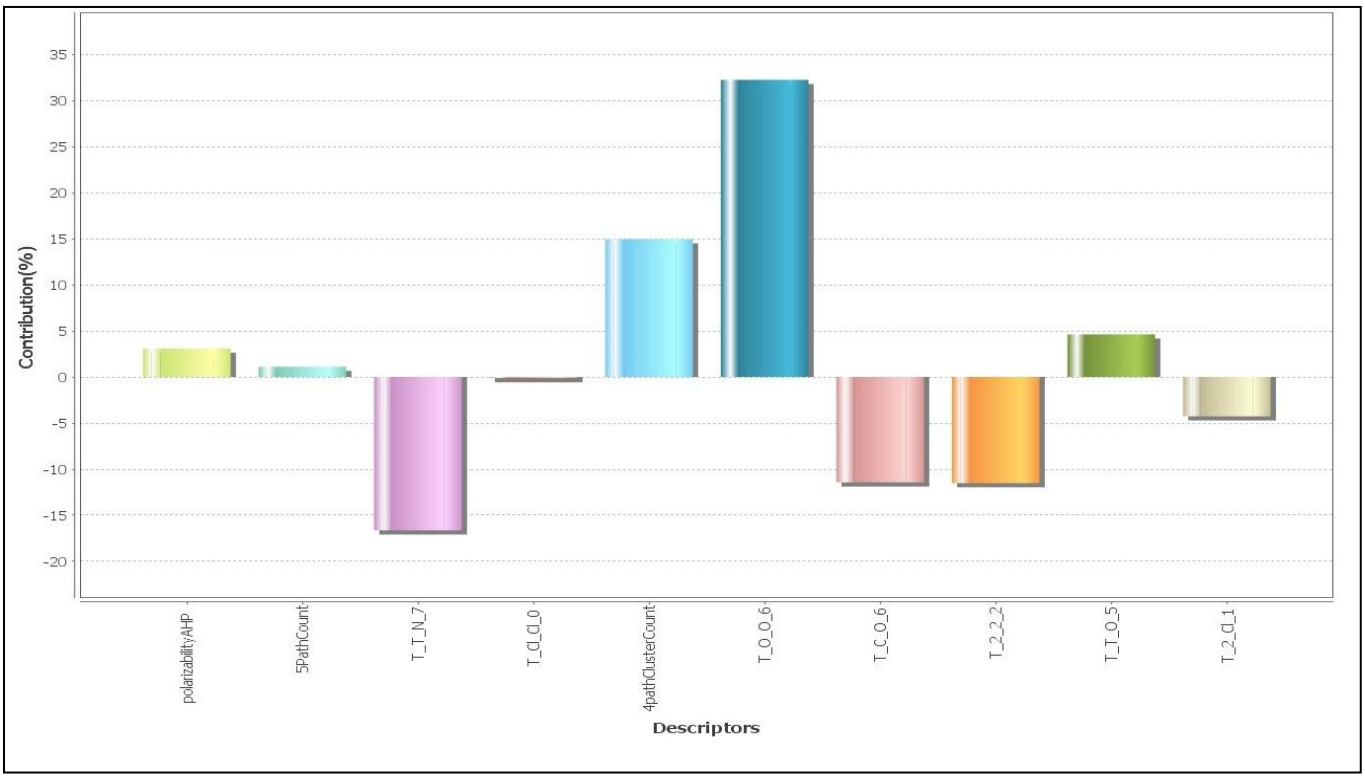

Figure 21: Contribution Plot of 2D QSAR Model 4

Table 9: Comparison of actual and predicted activities by the four 2D models

	ACTUAL ACTIVITY	**PREDICTED ACTIVITY**			
		MODEL 1 SVS-Multiple	MODEL 2 SVS-PLS	MODEL 3 SA-Multiple	MODEL 4 SA-PLS
OA1	8.064	8.476776	8.277796	8.021083	8.084438
OA10	8.737	8.840085	8.592003	8.94967	8.940079
OA11	8.082	8.24949	8.117734	8.224099	8.133052
OA12	9.004	8.86114	8.730217	8.946995	8.814196
OA13	8.26	8.600965	8.634371	8.436474	8.208163
OA14	8.069	8.050457	7.971693	8.054514	7.951035
OA15	8.477	8.59743	8.706183	8.516854	8.614481
OA16	8.006	8.779773	8.595796	8.484526	8.344126
OA17	8.115	8.560521	8.549969	7.851887	8.381672
OA18	9.832	9.725277	9.605908	9.977982	9.572596
OA19	8.499	8.288088	8.25126	8.395888	8.108822
OA2	8.368	8.240208	8.112658	8.535434	8.070947
OA20	9.86	9.641073	9.565545	9.891581	9.615666
OA21	9.271	9.686639	9.55188	9.361893	9.557436
OA22	8.876	9.197432	9.308563	8.709558	9.401059
OA23	7.795	7.596499	7.715855	7.546839	7.749213
OA24	9.991	9.513385	9.652873	10.153558	9.418901
OA25	10.058	9.236069	9.362591	9.324864	9.416218
OA26	10.089	9.855186	10.073222	9.973872	9.86731
OA27	9.516	9.928493	9.878163	9.588648	9.986058
OA28	10.178	10.375833	10.415754	10.135028	10.410219
OA29	10.669	10.296707	10.258167	10.560922	10.356189
OA3	10.55	8.499992	8.458874	7.49751	8.251962
OA30	8.576	8.818928	8.844243	8.900804	8.881285
OA4	7.756	7.655735	7.812062	7.875785	7.904668
OA5	8.804	8.488567	8.51209	7.83472	8.757123

OA6	8.947	8.567444	8.423695	8.905662	8.677034
OA7	7.329	7.362441	7.58247	7.449547	7.71957
OA8	8.37	8.285169	8.534082	8.423659	8.463124
OA9	9.341	8.840709	8.70239	9.148885	9.101206

Table 10: Statistical values of the 2D models

STATISTICAL VALUES	**MODEL1**	**MODEL2**	**MODEL 3**	**MODEL4**
N	23	23	23	23
Degree of freedom	17	19	12	18
r^2	0.9028	0.8746	0.9720	0.9066
q^2	0.7860	0.8047	0.6968	0.7292
F_test	31.5859	44.1556	41.6574	43.6887
r^2_se	0.3149	0.3384	0.2012	0.3000
q^2_se	0.4673	0.4223	0.6620	0.5108
pred_r^2	0.1483	0.2171	0.5794	0.1095
pred_r^2se	0.9990	0.9578	1.3604	1.0214

According to VlifeMDS set guidelines, QSAR models are significant if $r^2 > 0.7$, $q^2 > 0.5$ and pred_$r^2 > 0.5$. Thus comparing the statistical parameters of the four 2D QSAR models, we can consider the 3rd model to be statistically the best model.

4.4.2 ALIGNMENT OF MOLECULES

The molecules after molecular alignment were opened in VLifeMDS simultaneously.

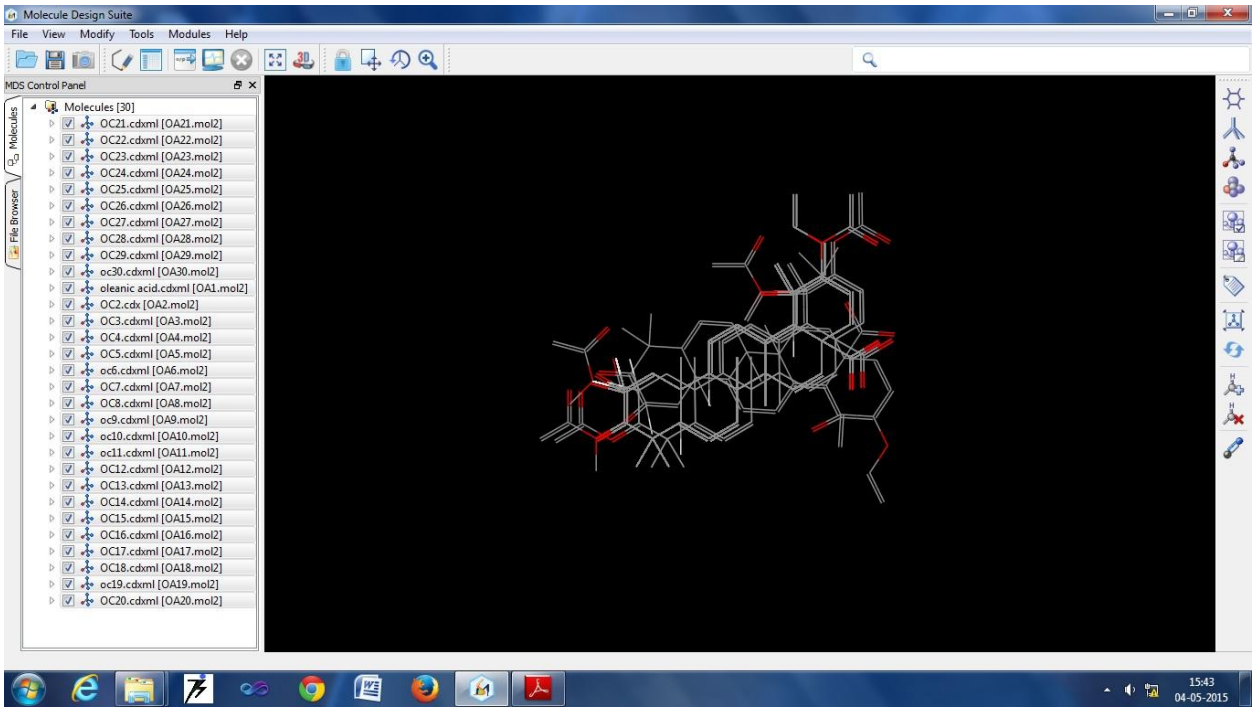

Figure 22: Aligned molecules w.r.t the skeleton structure.

4.4.3 3D QSAR

3D QSAR was performed using the "3D QSAR" module of VlifeMDS software. 3D descriptors were computed by calculating the electrostatic and steric molecular fields. Some 3D descriptor values and the Unicolumn statistics of the data set are tabulated below.

Table 11: Some 3D descriptor values

	ACTIVITY	E_1	E_2	E_3	S_1	S_2	S_3	S_4
OA1.mol2	8.064	-0.166	-0.236	-0.269	-0.011	-0.016	-0.018	-0.016
OA10.mol2	8.737	-0.139	-0.193	-0.218	-0.013	-0.019	-0.022	-0.019
OA11.mol2	8.082	-0.196	-0.274	-0.31	-0.012	-0.017	-0.019	-0.017
OA12.mol2	9.004	-0.156	-0.226	-0.259	-0.013	-0.019	-0.022	-0.019
OA13.mol2	8.26	-0.167	-0.24	-0.273	-0.011	-0.016	-0.019	-0.016
OA14.mol2	8.069	-0.193	-0.269	-0.303	-0.011	-0.016	-0.018	-0.016
OA15.mol2	8.477	-0.209	-0.287	-0.322	-0.011	-0.016	-0.018	-0.016
OA16.mol2	8.006	-0.184	-0.258	-0.292	-0.011	-0.016	-0.018	-0.016

OA17.mol2	8.115	-0.192	-0.268	-0.302	-0.011	-0.016	-0.018	-0.016
OA18.mol2	9.832	-0.216	-0.295	-0.331	-0.011	-0.016	-0.018	-0.016
OA19.mol2	8.499	-0.227	-0.307	-0.344	-0.011	-0.016	-0.018	-0.016
OA2.mol2	8.368	-0.238	-0.319	-0.356	-0.011	-0.016	-0.018	-0.016
OA20.mol2	9.86	-0.231	-0.312	-0.348	-0.011	-0.016	-0.018	-0.016
OA21.mol2	9.271	-0.212	-0.291	-0.327	-0.011	-0.016	-0.018	-0.016
OA22.mol2	8.876	-0.306	-0.413	-0.461	-0.012	-0.017	-0.019	-0.017
OA23.mol2	7.795	-0.265	-0.366	-0.412	-0.012	-0.017	-0.019	-0.017
OA24.mol2	9.991	-0.15	-0.219	-0.251	-0.012	-0.017	-0.02	-0.017
OA25.mol2	10.058	-0.271	-0.352	-0.388	-0.008	-0.011	-0.012	-0.011
OA26.mol2	10.089	-0.277	-0.38	-0.427	-0.012	-0.017	-0.019	-0.017
OA27.mol2	9.516	-0.172	-0.238	-0.269	-0.011	-0.016	-0.019	-0.016
OA28.mol2	10.178	-0.139	-0.201	-0.229	-0.011	-0.016	-0.019	-0.016
OA29.mol2	10.669	-0.162	-0.225	-0.254	-0.01	-0.014	-0.016	-0.014
OA3.mol2	10.55	0.101	0.15	0.174	-0.014	-0.022	-0.025	-0.022
OA30.mol2	8.576	-0.168	-0.24	-0.273	-0.011	-0.016	-0.019	-0.016
OA4.mol2	7.756	-0.266	-0.355	-0.396	-0.011	-0.017	-0.019	-0.017
OA5.mol2	8.804	0.065	0.1	0.119	-0.017	-0.026	-0.031	-0.026
OA6.mol2	8.947	-0.083	-0.105	-0.114	-0.014	-0.02	-0.024	-0.02
OA7.mol2	7.329	-0.145	-0.198	-0.221	-0.012	-0.017	-0.019	-0.017
OA8.mol2	8.37	-0.132	-0.18	-0.201	-0.011	-0.016	-0.018	-0.016
OA9.mol2	9.341	-0.225	-0.311	-0.35	-0.011	-0.016	-0.019	-0.016

Table 12: Unicolumn statistics

Column name	**Average**	**Maximum**	**Minimum**	**StdDev**	**Sum**
ACTIVITY (Training Set)	9.0171	10.6690	7.3290	0.9332	234.445
ACTIVITY (Test Set)	8.2610	8.4770	8.0820	0.1931	33.004

Descriptors that contribute to the model are S_349 and S_7. The range of S_349 is from -0.5905 to -0.5388 while the range of S_7 is from -0.0532 to -0.0346. A negative range of steric fields indicates that negative steric potential is favorable for an increase in the activity, and hence less bulky substituent group is preferred in that region.

k-Nearest Neighbour = 3, N_{test} = 4, $N_{training}$ = 26, Degree of freedom = 23, q^2 = 0.6548, q^2se = 0.5483, pred_r^2 = 0.8999, pred_r^2se = 0.2829

According to the model, the internal model validation predicts the activity of training set as 65% (q^2 = 0.6548) and external model validation predicts the activity of the test set as 90% (pred_r^2 = 0.8999).

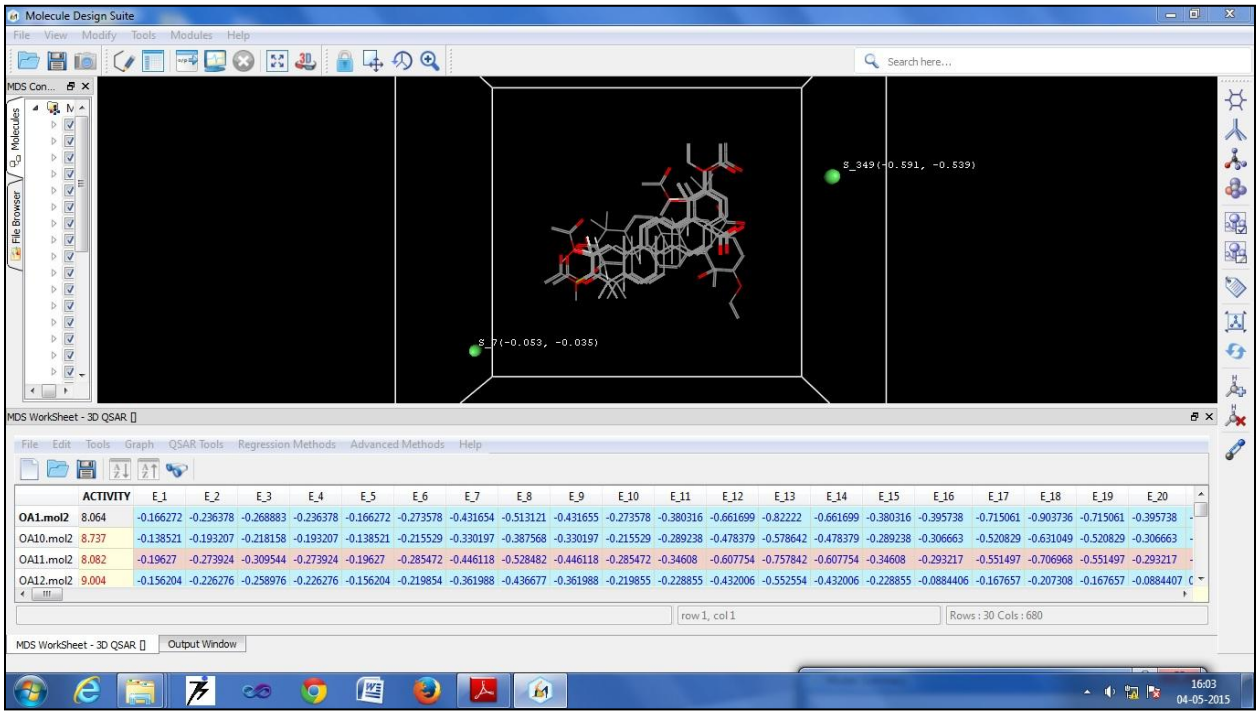

Figure 23: 3D QSAR model

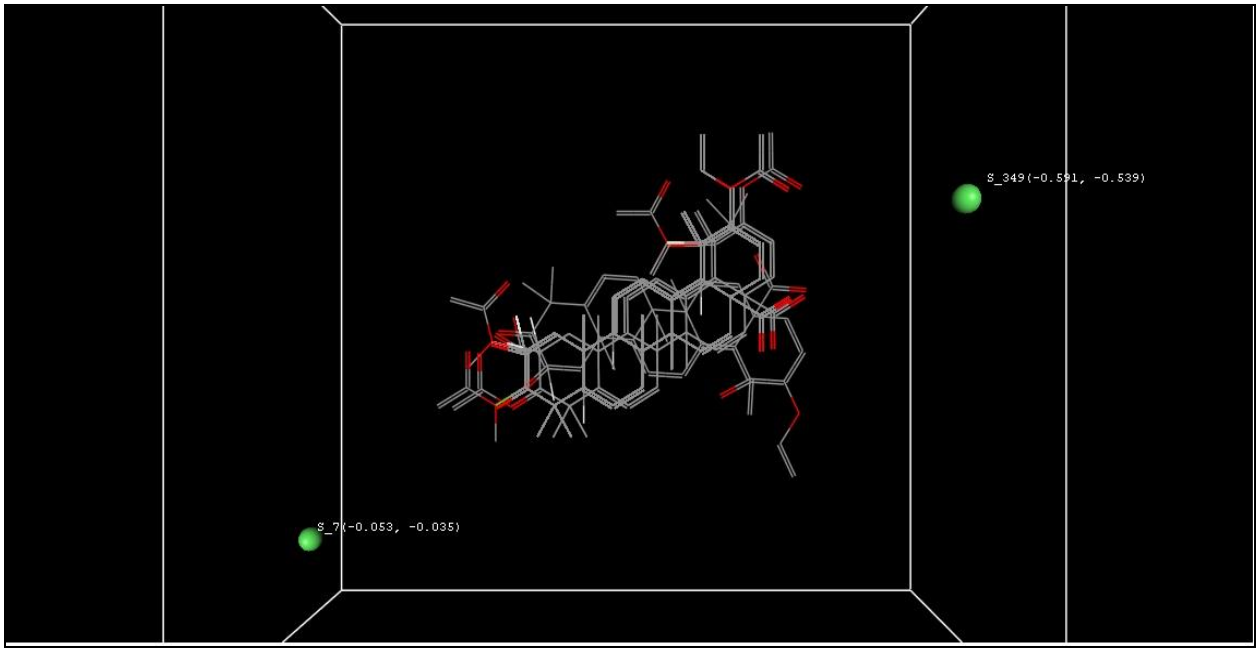

Figure 24: Steric molecular fields determined by the 3D model

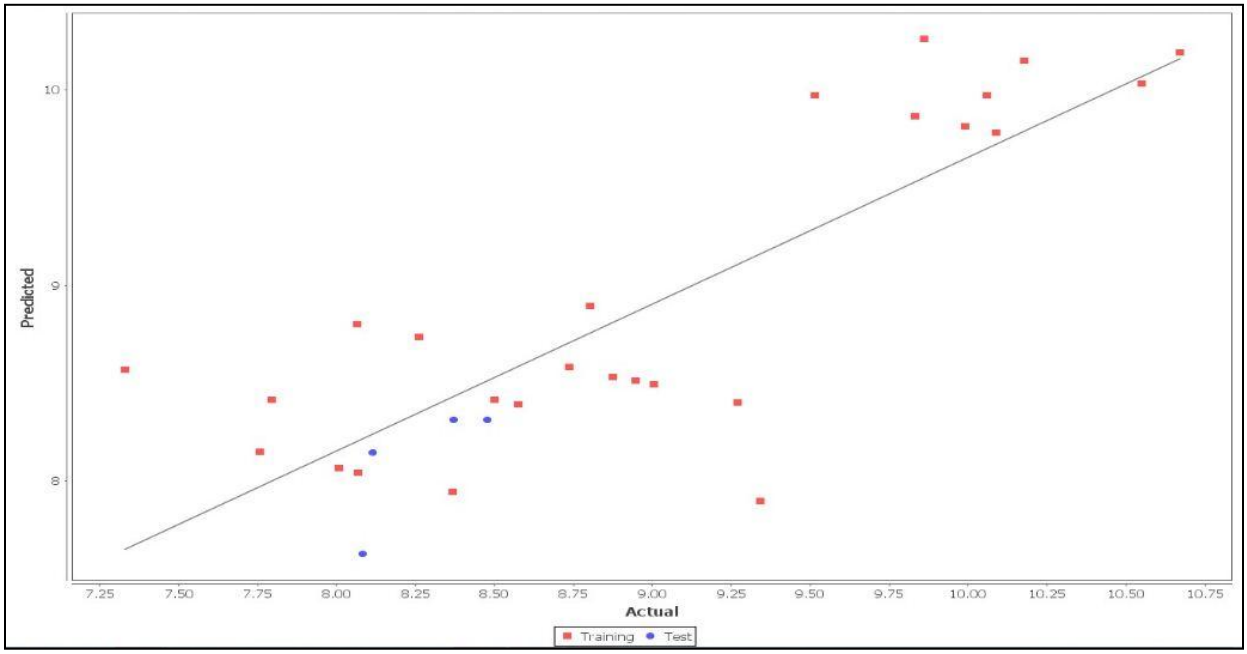

Figure 25: Fitness Plot of 3D Model

The activities of the molecules, as predicted by the 3D QSAR model, are as below –

Table 13: Comparison of actual and predicted activities by 3D QSAR model

MOLECULE	ACTUAL ACTIVITY	PREDICTED ACTIVITY
OA1	8.064	8.802181
OA10	8.737	8.583205
OA11	8.082	7.625713
OA12	9.004	8.494189
OA13	8.26	8.737086
OA14	8.069	8.043555
OA15	8.477	8.311277
OA16	8.006	8.064338
OA17	8.115	8.146974
OA18	9.832	9.864937
OA19	8.499	8.415672
OA2	8.368	7.943678
OA20	9.86	10.262446
OA21	9.271	8.399911
OA22	8.876	8.531852
OA23	7.795	8.414827
OA24	9.991	9.812584
OA25	10.058	9.972453
OA26	10.089	9.779635
OA27	9.516	9.970558
OA28	10.178	10.151561
OA29	10.669	10.189832
OA3	10.55	10.032097
OA30	8.576	8.389818
OA4	7.756	8.147806

OA5	8.804	8.896068
OA6	8.947	8.513603
OA7	7.329	8.570383
OA8	8.37	8.31226
OA9	9.341	7.899583

CHAPTER 5:

CONCLUSION

CONCLUSIONS:

In our current study, we tried to narrow down our search for modulators that can promote the effect of NGF by binding to its receptor, and thus be used as therapeutic tools to supplement enhanced nerve regeneration. Based on our docking results, the best docked ligand was kaempferol-3-glucoside. However, due to its poor ADME properties, we performed further studies with the next set of molecules having equally good binding affinity with the receptor. With binding energy of approximately -9 kcal/mol, all these molecules were found to have a structure similar to triterpenoids. These molecules were found to be oleanolic acid, taraxerone, taraxerol, alpha and beta amyrin. These molecules were also found to have suitable ADME properties to be considered as candidate drug molecules. We further generated mathematical models using 2D and 3D Quantitative Structure-Activity Relationship(QSAR) to establish a relationship between the structure and activity. For this, we chose oleanolic acid as the skeleton structure and its derivatives were designed. The 2D and 3D models generated, provide us an efficient way to predict the activity of any new analog with similar structures or design a new molecule with better activity.

REFERENCES:

1. Christine E. Schmidt and Jennie Baier Leach. **Neural Tissue Engineering: Strategies for Repair and Regeneration.** Annu. Rev. Biomed. Eng. 2003. 5:293–347
2. Ai J, Kiasat-Dolatabadi A, Ebrahimi- Barough S, Ai A, Lotfibakhshaiesh N, Norouzi-Javidan A, et al. **Polymeric Scaffolds in Neural Tissue Engineering: A Review**. Arch Neuro Sci. 2013; 1(1): 15-20
3. Stephanie M. Willerth and Shelly E. Sakiyama-Elbert. **Approaches to Neural Tissue Engineering Using Scaffolds for Drug Delivery**. Adv Drug Deliv Rev. 2007; 59(4-5): 325–338.
4. Tessa Gordon. **The role of neurotrophic factors in nerve regeneration**. Neurosurg Focus 2009; 26.2.E3
5. Ella Bossy Wetzel, Robert Schwazenbacher and Stuart A Lipton. **Molecular pathways to neurodegeneration**. Review Article- Nature Medicine
6. C. A. Davie. **A review of Parkinson's disease**. British Medical Bulletin 2008; 86: 109–127
7. J Jankovic. **Parkinson's disease: clinical features and diagnosis**. J Neurol Neurosurg Psychiatry 2008;79:368-376
8. Ernfors P., Ibáñez C. F., Ebendal T., Olson L. and Persson H. **Molecular cloning and neurotrophic activities of a protein with structural similarities to nerve growth factor: developmental and topographical expression in the brain**. Proc. Natl. Acad. Sci. USA, 1990; 187**:** 5454–5458
9. Hallböök F., Ibáñez C. F. and Persson H. **Evolutionary studies of the nerve growth factor family reveal a novel member abllndantly expressed in Xenopus ovary**. Neuron **6,** (1991) 845–858
10. Götz R., Köster R., Winkler C., Raulf F., Lottspeich F., Schartl M. et al. **Neurotrophin-6 is a new member of the nerve growth factor family**. Nature 1994; 372: 266–269
11. Thomas P. Misko, Monte J. Radeke and Eric M. Shooter. **Nerve growth factor in neuronal development and maintenance.** Biol.1987; 132, 177-190
12. C.Wiesmanna and A. M. de Vos**. Nerve growth factor: structure and function.** Cell. Mol. Life Sci. 58 (2001) 748–759

13. Luigi Aloe**,** Maria Luisa Rocco**,** Patrizia Bianchi **and** Luigi Manni**. Nerve growth factor: from the early discoveries to the potential clinical use.** Journal of Translational Medicine 2012, 10:239

14. Allen SJ, Dawbarn D. **Clinical relevance of the neurotrophins and their receptors**. Clin Sci (Lond) 2006, 110:175–191
15. Bothwell M. **Functional interactions of neurotrophins and neurotrophin receptors**. Annu. Rev. Neurosci. **18:** (1995) 223–253
16. Yano H. and Chao M. **Neurotrophin receptor structure and interactions**. Pharm. Acta. Helv. (2000) **74:** 253–260
17. Cordon-Cado C., Tapley P., Jing S. Q., Nanduri V., O'Rourke E., Lamballe F. et al. **The trk tyrosine protein kinase mediates the mitogenic properties of nerve growth factor and neurotrophin-3**. Cell (1991) **66:** 173–183
18. Klein R., Nanduri V., Jing S. A., Lamballe F., Tapley P., Bryant S. et al. **The trkB tyrosine protein kinase is a receptor for brain-derived neurotrophic factor and neurotrophin-3**. Cell (1991) **66:** 395–403
19. Jing S, Tapley P, Barbacid M. **Nerve growth factor mediates signal transduction through trk homodimer receptors.** Neuron 1992; 9(6):1067-79
20. **Alzheimer's Association Report**. Alzheimer's & Dementia 11 (2015) 332–384
21. Ella Bossy-Wetzel, Robert Schwarzenbacher and Stuart A Lipton. **Molecular Pathways to neurodegeneration**. Nature Medicine (2004); Vol 10 pgs S2-S9.
22. Selkoe D.J**. Alzheimer's disease is a synaptic failure**. Science 298 (2002); 789-791
23. Hirai K. **Mitochondrial abnormalities in Alzheimers's Disease**. J. Neuroscience 21 (2001); 3017-3023.
24. Kamal A, Stokin GB, Yang Z, Xia CH, Goldstein LS. **Axonal transport of amyloid precursor protein is mediated by direct binding to the kinesin light chain subunit of kinesin-I**. Neuron (2000);28(2):449-59
25. C. A. Davie. **A review of Parkinson's disease.** British Medical Bulletin 2008; 86: 109–127
26. Mark S Forman, John Q Trojanowski & Virginia M-Y Lee. **Neurodegenerative diseases: a decade of discoveries paves the way for therapeutic breakthroughs**. Nature Medicine **10** (2004), 1055 - 1063

27. April M Weissmiller and Chengbiao Wu**. Current advances in using neurotrophic factors to treat neurodegenerative disorders.** Translational Neurodegeneration 2012, 2047-9158-1-14.
28. Calissano P, Matrone C, Amadoro G. **Nerve growth factor as a paradigm of neurotrophins related to Alzheimer's disease**. Dev Neurobiol 2010, 70:372–383.
29. Jonathan D. Cooper, Ahmad Salehi, Jean-Dominique Delcroix, Charles L. Howe, Pavel V. Belichenko, Elaine J. Carlson§, Charles J. Epstein§, and William C. Mobley. **Failed retrograde transport of NGF in a mouse model of Down's syndrome: Reversal of cholinergic neurodegenerative phenotypes following NGF infusion**. PNAS 2001; vol. 98, 10439–10444.
30. Haitao Wang, Rikang Wang, Thilini Thrimawithana, Peter J. Little, Jiangping Xu, Zhong-Ping Feng, and Wenhua Zheng. **The Nerve Growth Factor Signaling and Its Potential as Therapeutic Target for Glaucoma**. BioMed Research International 2014, Article ID 759473, 10 pages
31. Sima AA. **New insights into the metabolic and molecular basis for diabetic neuropathy.** Cell Mol Life Sci. 2003; 60(11):2445-64.
32. Apfel SC, Arezzo JC, Brownlee M, Federoff H, and Kessler JA. **Nerve growth factor administration protects against experimental diabetic sensory neuropathy**. Brain Res.1994 ;634(1):7-12
33. Carlo Pincelli, Cinzia Sevignani, Rossella Manfredini, Alexis Grande, Fabrizio Fantini, Luisa Bracci-Laudiero, Luigi Aloe, Sergio Ferrari, Andrea Cossarizza and Alberto Giannetti**. Expression and Function of Nerve Growth Factor and Nerve Growth Factor Receptor on Cultured Keratinocytes**. Journal of Investigative Dermatology 1994; **103**, 13–18
34. Mukesh C. Sharma. **Molecular Modelling Studies for the Discovery of New Substituted Pyridines Derivatives with Angiotensin II AT**1 **Receptor Antagonists.** Interdiscip Sci Comput Life Sci 2014; 6: 197–207
35. Mukesh C. Sharma, Smita Sharma and K. S. Bhadoriya. **Molecular modeling studies on substituted aminopyrimidines derivatives as potential antimalarial compounds**. Med Chem Res 2013; 014-1199-2.

36. Sanmati Kumar Jain, Ravi Tripathi and Piyush Ghode. **Conventional QSAR studies of thiazolidinone derivatives as potential epidermal growth factor receptor inhibitors.** World Journal Of Pharmacy And Pharmaceutical Sciences 2014; Volume 3, 1811-1821.